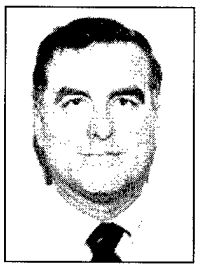
The Bulletin of Military Operations Research

PHAL

Vol. 29 No. 3 (ISSN 0195-1920) September 1996

Model Verification, Validation, and Accreditation (VV&A) Common Ground Within the M&S Community

Dr. Patricia Sanders
OSD/DDTSE&E



Ray Miller
AF/XOMT

For the past two years, the Department of Defense has placed increased emphasis on both assessing and enhancing the credibility of its models and simulations (M&S) through the development of practical, cost effective VV&A methods and techniques, as well as seeking the tools needed to automate these efforts. This work is designed to be extensible across the spectrum of DoD M&S that support various analytic, training, wargaming/exercise, and acquisition activities. These VV&A methodologies must address problems and issues currently facing our DoD practitioners such as:

- Data availabilty, quality, and standardization;
- V&V sufficiency (How much V&V is enough?);
- Appropriate V&V methods for standalone or distributed model suites (or federations) used for specific domain applications;
- Appropriate method/level of V&V for the large number of existing (legacy) systems that will be used until their successor(s) becomes available; and, most importantly
- What are the costs—and benefits—realized from VV&A?

The Defense Modeling and Simulation Office (DMSO)-sponsored VV&A Technical Working Group (TWG) serves as the DoD forum which identifies these VV&A problem areas and draws together the appropriate expertise to determine possible approaches to these issues. Many of these issues are breaking new ground within DoD; however, have these issues been addressed by M&S organizations outside of DoD? The TWG members felt that a conference comparing the VV&A challenges, methods, tools, and results experienced by the various M&S organizations would bring fresh perspectives to the DoD effort. Where do we share common ground of what VV&A is or should be? Where do our experiences diverge—and why? More importantly, what methods and tools are found to be most cost-effective for particular modeling applications?

Subsequently, a Colloquium involving M&S experts from government, industry, and academia was conducted at Fort Belvoir, VA on 28 and 29 Sep 95. The agenda consisted of a series of presentations by practioners who are actively engaged in their organizations' V&V efforts. Two presentations specifically addressed software V&V activities; the remaining presentations were case studies illustrating how different organizations approached model V&V. (An Executive Summary and briefing slides for the each presentation, as well as an expanded white paper, is available for download from the DMSO Home Page — http://www.dmso. mil.). These presentations included:

- Mr. **George Neat,** U.S. Department of Transportation, *Crack Failure Models for Aircraft;*
- Dr. **Peter Knepell** and Dr. **Thomas Curry,** Logicon RDA, *VV&A for an Acquired Immune Deficiency Syndrome (AIDS) Model;*
- Mr. **John Hinkle,** NASA Software IV&V Facility, *IV&V of the International Space Station* (Software);
- Mr **Brian Keltch,** National Institute for Petroleum and Energy Research, *Department of Energy's Tertiary Oil Recovery Information System VV&A Issues;*
- Mr. **Bruce Hicks,** NOAA Air Resources Laboratory, *(Atmospheric) Dispersion Modeling;*
- Dr. **Lance Miller,** SAIC, *Quality Software: Lessons Learned*
- Mr. **Jeff Gordon,** U.S. Department of Transportation, *Lessons Learned Using Finite Element and Non-Linear Optimization Models to Predict Railroad Component Stress and Performance in Service;* and
- Dr. **Michael Kleinberger,** National Highway Traffic Safety Administration, *Development and Validation of Human Anatomic Finite Element Models;*

These presentations served as both a

(See **VV&A** *p. 30)*

THE ISSUE TODAY

Leaders' Notes

Features

Departments

Military Operations Research Society
101 S. Whiting Street, #202
Alexandria, VA 22304
(703) 751-7290; FAX (703) 751-8171
e-mail: morsoffice@aol.com
http://www.msosa.mil.inter.net/mors/

The *PHALANX* (ISSN 0195-1920) is published quarterly, $20.00 for one year or $35.00 for two years (US Rates) by the Military Operations Research Society in cooperation with the Military Applications Society (MAS) of the Institute for Operations Research and Management Science (INFORMS). Principal office: Landmark Towers, 101 S. Whiting Street, #202, Alexandria, Virginia 22304. Second-class postage paid at Alexandria, and at additional mailing offices. POSTMASTER: Send address changes to *PHALANX*, 101 S. Whiting St., #202, Alexandria, VA 22304.

INFORMS
901 Elkridge Landing Road, Suite 400
Linthicum, MD 21090
(410) 850-0300 (800) 446-3676
FAX (410) 684-2963
e-mail: informs@jhuvms.hcf.jhu.edu

Subscriptions to *PHALANX* are included in the annual dues of INFORMS/MAS members or meeting registration fees for attendees at MORS sponsored meetings. Others may subscribe for $20 for one year or $35 for two years (US Rates).

Printed On Recycled Paper

MORS PRESIDENT

MORS 1996-97 a Year for "Reestablishing the Foundations of Analysis"

Fred Hartman
MORS President

In the June issue of *PHALANX* you read the "Platform" articles prepared by each of the Candidates running to be your new MORS President. We had each prepared very logically similar articles in isolation that were each centered on the theme of Challenge, Change and Opportunity. Not having seen the articles prepared by **Priscilla Glasow** and **Jackie Henningsen**, I was very surprised upon picking up my copy of the *PHALANX* to find that even our titles were very similar. In keeping with these thoughts and my platform statements, this will be a MORS year of leveraging technology, managing and controlling change and preparing our Society and Profession for the next (some would even say current) era of Joint, Interoperability and Coalition Operations. Our community must be prepared to accommodate the explosive growth in information/communication technologies and take advantage of that opportunity for creating the proper, up to date tools for our analysis profession with applications in this new environment.

Your new slate of elected Society Officers also includes **Jerry Kotchka**, VP (Finance and Management); **Kerry Kelley**, VP (Meeting Operations); **Denny Baer**, VP(Professional Affairs); and **Joe Tatman**, Secretary. This is an incredibly capable and experienced team that has spent many years in MORS positions even before their present term on the Board of Directors. In our formative, planning meetings at Fort Leavenworth we came to rapid agreement on the need for wider participation and empowerment of our membership at large and will be soliciting not only your continuing, but increasing levels of support and activity for the Society. We have been involved in adding Non-Board Members to each standing MORS committee and are always looking for even more volunteers. We will be soliciting your help very soon in our effort to "reengineer" the Working Group / Composite Group Structure. A special committee was formed under Kerry Kelley, chaired by **Sue Iwanski** which will be laying out their plan of action to accomplish the details this year for implementation at the 66th MORSS (the MORSS after next) in 1998.

The above MORS Theme for this year is a consensus product of your officer team. As we labored last month over the articulation of the Society's thrust for the coming year, these simple words, in our mind, expressed what we are about in the Society for the next year.

Looking Back

Our immediate Past President, **Chris Fossett** has led the Society in a very warm and efficient manner over the past year and presents me with a real challenge to keep up with her effective style of leadership. I'm most thankful for the excellent shape in which she has handed over your Society to our new Officer Team – and that she and our Advisory Directors (AD) are available to guide our steps during the coming months. The slate of AD's who very graciously agreed to continue their active involvement with the Board of Directors represents a dedication to MORS and our professional community that makes this organization so very special.

The 64th MORSS is now history! What a great Symposium!! We have much to be proud of in the success of our 64th at Fort Leavenworth. Our Program Chair, **Dick Helmuth** and his team pulled off a first class Symposium complete with some of the best site coordination work that we have ever seen!! A big special "Thank You" to Dick, **Mike Bauman**, **Phil Kubler**, **Annie Patenaude**, **Stuart Starr**, Kerry Kelley and the entire Program Committee for a job very well done. We are all most appreciative for a professional job that produced a "warm" Kansas atmosphere for both a productive and most enjoyable Symposium. Take the time to tell the 64th Team to include our MORS Staff how much we appreciate their hard work. **Dick Wiles** tells me that the final number stands at 920 which ranks as the 4th MORSS in terms of attendance. Not bad for Kansas in August – er June!

Wanner Award: Congratulations again to our MORS friend, Fellow, Past President and Chair of numerous MORS Meetings, **Ed Brady** on his selection for the Wanner Award. Ed becomes the 19th Wanner Award Laureate!

Looking Forward

After collectively basking in the glow of our great past year and Symposium, it is time to roll up our sleeves and start to work on the next great MORS year. We have a tremendous set of Vice Presidents and Committee Chairs along with a large slate of Non-Board Members who have volunteered to help with our Committees.

Congratulations to your following new members of the Board of Directors which were elected to their standard four year term at our June Meeting: Col **Tom Allen**, Air Force Studies and Analysis Agency; Dr **Al Brandstein**, Marine Corps Combat Development Command; RADM **Pierce Johnson**, Naval Reserve Readiness Command, Region 6; MAJ **Willie McFadden**, Army, HQ TRADOC; Dr **Julian Palmore**, University of Illinois, and Editor of *PHALANX*; and Dr **Roy Rice**, Teledyne Brown Engineering. We are proud to have such accomplished and widely diversified Morsians joining the Board of Directors and expect great things out of them over their next years of service. This year there were 18 outstanding nominations voted on to fill the above six slots.

The 65th MORSS is off to a good start with **Harry Thie**, Program Chair conducting kick off meetings with his excellent Program Committee at Fort Leavenworth. He is out in front with a superb committee that is deep in talent and long on reputations for hard work. It is most impressive to see the degree of organization that is already in place for next summer. Let's keep at it!! Harry has put a detailed "Planning Schedule" in place that has some very important work to be done over the next few months. Please be thinking early about the 65th MORSS theme of "Analysis for Complex and Uncertain Times". We need to incorporate that thought into everything we do this year. This theme also fits very well with the

(See **MORS PRESIDENT** *p. 29)*

Recognizing Excellence

Dr. Stephen J. Balut
MAS President

Our members are making important contributions to strengthening national security. We do this by providing information that improves defense decision making. I hope you will join me in congratulating a few of our fellow members who have recently been recognized for outstanding accomplishments as researchers and students of operations research.

Koopman Prize

At the Washington, DC INFORMS meeting, I was pleased to watch Professor **Keith Womer,** MAS's Prize Committee Chair, award the Koopman Prize for the outstanding military operations research study published in 1994 to Dr. **Hanif D. Sherali**, Virginia Polytechnic Institute and State University, Dr. **Youngho Lee**, U. S. West Advanced Technologies, and Dr. **Donald D. Boyer**, Naval Surface Warfare Center. Their winning study, "Scheduling Target Illuminators in Naval Battle-Group Anti-Air Warfare" was published in Naval Research Logistics. This outstanding effort was judged the superior piece among an unusually strong field of candidate papers. The papers were evaluated by the Koopman Prize Committee consisting of Professor **Jeff Camm**, University of Cincinnati, Dr. **Bob Tripp**, Synergy, Inc., and Professor **Don Barr**, U. S. Military Academy.

The Koopman Prize is named for the late **Bernard Koopman** (1900-1981), a pioneer in the field of operations research who was active in the founding of the Operations Research Society of America (ORSA) and served as its president in 1956. Koopman served as an operations research liaison between the U. S. Department of Defense, the U. K. military establishment and NATO, and played a key role in introducing operations research as a permanent NATO activity.

Awards for Excellence in Operations Research

One of MAS' most important missions is to encourage students of military operations research. We do that by fostering Student Chapters at military academies and universities, providing speakers, encouraging participation in INFORMS activities, and most importantly for today's message, by presenting awards to the outstanding operations research students at the military academies.

I was proud and honored to represent MAS at graduation ceremonies at the Naval Academy and the Air Force Academy. In Colorado Springs, I presented a MAS prize plaque and bronze eagle to Air Force Cadet **Andrew E. Coop** who was unanimously selected by faculty representatives from four Academic Departments. Cadet Coop was number one academically in his class in the Operations Research major and was also named Outstanding Squadron Commander. He was president of the Omega Rho Operations Research Honor Society at the Air Force Academy.

I attended a graduation award ceremony in Annapolis where I presented a MAS prize plaque and a check for $250 to Midshipman **Terrence Nawara** who excelled in courses in operations research. His selection was also based on a senior project, titled "Problems with Multi-line Sonar Detection," that was initiated during an internship at the Institute for Defense Analyses. Midshipman Nawara was selected for the nuclear power, subsurface warfare specialty.

Cadet **Tyler B. Smith** was awarded a MAS prize plaque and $250 worth of books at the Awards Convocation for the graduating class of 1996 at West Point. Cadet Smith earned the highest GPA in 10 core operations research courses. The award was presented by Colonel **David C. Arney**, Head of the Mathematical Sciences Department, U. S. Military Academy. I want to thank Professor **Don Barr** for assisting with the arrangements for the award of this prize.

Other News

In other Society news, I officially appointed the Nominating Committee for our upcoming election of officers. The Chair of the Committee is our President-elect, Professor **Tom Gulledge**. He is accompanied on the Committee by Dr. **Dean Hartley** and Dr. **Peter Cherry**, our two most recent past MAS Chairs. Watch for the announcement of nominees in this column.

Over the past few weeks, the INFORMS Business Office discovered that a large portion of MAS members were not receiving *PHALANX*. The root cause of this has been identified and is related to shifts in administrative responsibilities that followed the merger of ORSA and TIMS. A procedure has been reestablished and implemented that ensures that all members will receive each issue. To those of you who missed back issues, we are very sorry for this and we will be in touch with you.

I am pleased to report that membership in MAS is increasing; Clusters planned for Atlanta and San Diego look great; we are rewriting our Bylaws to reflect Society status; and we're negotiating to make it easier and cheaper for you to subscribe to our journal, *Military Operations Research*. Its a good time to get more actively involved. ✪

"Meetings: What's Ahead"

Kerry Kelley
VP (Meeting Operations)

"Enhance the quality and usefulness of classified and unclassified military operations research" – the stated purpose of MORS serves to underscore the goals and objectives of the meetings sponsored by our society. In reflecting upon our society's most recent meetings and in looking ahead to plans for the coming year it is useful to check our accomplishments and goals with the broadly stated purpose of our MORS. In doing so, I believe that you will find, as I did, a myriad of activities within meeting operations which help further this purpose. During the next year, I am charged with the Meeting Operations business of MORS. As the meeting operations responsibilities have transitioned to me, I find that I have been given an extremely orderly state of affairs – thanks to the strong leadership of **Fred Hartman**, who was VPMO last year and our President this year, **Chris Fossett**, our Past President, and the entire MORS staff. In the next few paragraphs I hope to update you on recent accomplishments and inform you as to what to expect in the coming year.

64th MORSS

The 64th MORSS in Ft. Leavenworth was an overwhelming success. We owe many thanks to the Program Chair, **Dick Helmuth**, and to his entire Program Committee, with special thanks to **Mike Bauman, Phil Kubler**, **Annie Patenaude** and **Stuart Starr**. The latest tally on attendance indicated that we were 920 members strong – the 4th largest MORSS attendance figure. I attribute this mostly to the outstanding program but I also cannot overlook the fine location as well (being a midwesterner myself). I hope all who participated found the symposium useful and are actively involved in leveraging technology in their military analyses.

65th MORSS

The 65th MORSS will be at the Marine Corps University in Quantico, VA on 10-12 June 1997. The theme will be "Analysis for Complex, Uncertain Times". Dr. **Harry Thie**, the Program Chair, is already working diligently at the preparations for this meeting and has also written an article in this edition of the *PHALANX*. Please see his article for additional information and for the points of contact on his 65th MORSS committee. **Jay Wilmeth**, the Working Group/Composite Group Coordinator and a member of the 65th MORSS committee will be working with the WG/CG Chairs, Co-chairs and Advisors to ensure a challenging agenda for all.

66th MORSS

Believe it or not, plans are already well underway for the 66th MORSS. The 66th MORSS will be hosted by the Naval Postgraduate School in Monterey, CA. We are fortunate to have RADM **Pierce Johnson** as the Program Chair for the 66th MORSS. With Pierce Johnson at the helm and the very popular NPS location, the 66th MORSS promises to be yet another outstanding conference.

Special Meetings

Dr. **Stuart Starr** has once again agreed to take the lead as the Chair of the Special Meetings Committee. He served MORS excellently in this capacity last year and under his leadership we benefited from several meetings:

- Developing a Framework for Joint Mobility Analysis (26-28 Sep 95)
- Advanced Distributed Simulations for Analysis 1996 (ADSA-1996) (30 Jan-1 Feb 96).

The workshop on joint mobility analysis was chaired by Mr. **Jim Johnson**, OASD(PAE) and advanced military OR through development of a framework for the analysis of mobility issues. The focus of ADSA-1996 was on the appropriate use of ADS, identification of advantages, limitations and areas for which improvements offer greatest potential, and suggested ways to analyze and test the quality of ADS-based tools. The workshop was chaired by MORS Fellow and Wanner Award Laureate, **Ed Brady** — an article appears in this issue of the *PHALANX*, page 13.

Stuart Starr has preparations well in hand for a mini-symposium on 1-3 October 1996 in McLean, VA entitled "First Order Analysis: Exploring Quick Response Analysis Requirements and Methodologies (QRAM). Dr. **Jackie Henningsen** is the General Chair; the Technical Chair is **Doug Williams** and Co-chairs are Dr. **Roy Rice** and **Bob Statz.** The meeting will provide a focus on existing methodologies that allow analysts to gain first order understanding of problems prior to allocation of analytic resources and to provide quick response results for senior DoD leaders.

In the planning stages are preparations for two additional special meetings:

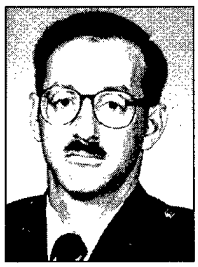
- "State of the Art in Warfare Analysis" (late 1996 - Chairs: **Gene Visco,** FS and Dr. **Al Brandstein**)
- "Modeling and Simulation (M&S) Linkage" (Spring 1997 - Chairs: Dr. **Marion Williams,** FS and **Jim Sikora,** FS)

The objective of the meeting on "State of the Art in Warfare Analysis" is to build on a recent international workshop on analytic approaches for the study of future conflict. The "M&S Linkage" mini-symposium is intended to address the state-of-the-art in M&S linkage and current issues and concerns for this exploding trend. Look for additional information in upcoming *PHALANX* issues.

One final item which we are exploring is the possibility for MORS to organize an International Symposium for selected allies' participation. Our research is only in its infancy at this point but the concept was suggested by the MORS Sponsors and Past Presidents and we will investigate this very exciting potential!

(See **VEEPS PEEP** *p. 35*

Objects, Objections, and Models



Major Kirk A. Yost
Naval Postgraduate School

Introduction

In a previous article in *PHALANX*, I took an abbreviated shot at the current enthusiasm for huge, objected-oriented, distributed interactive simulations ("Measuring Pants Legs," December 1995). Both my current research and the recent lead article in *PHALANX* ("Simulation for C^{4}ISR," by **Chuck Marshall** and Dr. **Randy Garrett**, March 1996) now permit me to attack on a broader front about the viability of these fashions in analysis. I'm not acquainted with Dr. Garrett, but Chuck Marshall is a former colleague of mine and I fully expect him to respond (read shoot back).

For the last few years, I've been helping develop an optimization model for munitions planning and campaign analysis, and my patrons recently asked me to add C^{4}ISR to this model in some form. Marshall and Garrett's article appeared at about the same time, so I'd hoped it would provide some insight as to what's going on in the mainstream with C^{4}ISR. Marshall and Garrett's piece did give me a great deal of insight, but not the type I was seeking. As a result, I offer this short piece as both a response to their claims and a reminder to the community that fashionable techniques and useful analyses are not the same.

Their Objects and My Objections

Marshall and Garrett are clearly enamored of objects. They call for a new class of simulations to be built entirely at the entity level, and even suggest using objects to represent things such as tactics and doctrine. Such an approach, they assert, will make life better, because current aggregate models rely on statistical modeling of collections of objects, and this clearly can't be as good as modeling the objects themselves. In addition, they claim an object approach is more computable, because "the computational load can be distributed among multiple processors, located at various sites."

Marshall and Garrett reveal a lot in their account for why C^{4}ISR is hard to model: the primary reason, they say, has been the lack of computational power, while achieving representation of human decision making and communications has been *secondary*. This proposition is amazing enough that I will repeat it: our biggest problem modeling C^{4}ISR is the lack of computational capability; the fact that we don't know how to model the essential elements and their interactions isn't that big a deal. The authors are convinced that advances in computing and the switch to object orientation will overwhelm these secondary problems, and somehow solve the problems of modeling C^{4}ISR. I contend this argument is wrong, and is central to a larger debate about object-oriented methods.

Under the subheading "Modeling Information as an Object," the authors claim that "[traditionally] in order to accommodate the contribution of intelligence, stochastic relationships were developed between objects. As a result, the value of intelligence is difficult to ascertain and its relationships and other factors on the battlefield are never determined." It appears that this statement points to a collection of bad models and rejects analytical approaches based on those models' behavior. Marshall and Garrett would no doubt include the optimization I'm working on, as it has no C^{4}ISR whatsoever (except for rudimentary BDA), and no objects of any kind. Now, this model's current shortcomings include not addressing C^{4}ISR, but to say these shortcomings are due to its lack of "real, tangible, and visible" C^{4}ISR objects is insupportable: again, C^{4}ISR is simply not modeled. Nonetheless, Marshall and Garrett make just this argument when they say later that "the current structure of most models does not lend itself to illuminating the worth of information." They haven't proved that analytical methods are incapable of representing C^{4}ISR; they have only stated that some analytical models do C^{4}ISR badly. Does that say anything conclusive about what analytical approaches could do if someone bothered to invest in them?

Consider two phenomena that are currently modeled using analytical methods: first, the movement of hurricanes; and second, the inflation factors used by the government for various policy decisions. Both are tremendously complex and governed by processes that we don't understand well (like C^{4}ISR). However, we have used these models for a long time and enjoyed a fair amount of accuracy. If we attempted to model either of these phenomena using the molecular approach advocated by Marshall and Garrett, how long would it take us to even *develop* a model? Don't be fooled by romantic terms such as object request brokers, inheritance, and polymorphism — any model developed under these philosophies will be *very* large and *very* complicated.

Even if I was willing to accept object modeling at the level proposed by Marshall and Garrett, I would balk at the greatest gap in their approach: the demotion of the machine representation of human decision making to a second-string problem. We know human logic is extremely variable and human choices and priorities are generally not transitive; the AI community continues to struggle with representing even the simplest decision using machine logic. However, to merely assure us that "work on higher-level 'strategic thinking' is on the way," grossly overstates our current or near-term capabilities. At the most recent INFORMS meeting in Washington, D.C., I attended a military applications seminar that was discussing AI approaches to command in models. One speaker was covering the capabilities of his organization's ground combat model, which contained a very complex AI-based command system. At the end, I asked what it took to build a rule set for this model. The speaker replied he thought it would take three Army officers with both command experience and an MS or Ph.D. in artificial intelligence about a year to do the job, *for a particular scenario*. How responsive is this?

Objects, Combinatorics, and the Lessons of Regression

The primary danger of using object-oriented methods is that they are combinatorially explosive. Marshall and Garrett, while bemoaning the lack of explicit sensor modeling, say that "sensors want to see 'things,' not an equation representing thousands of things." Let's consider the ramifications of stuffing thousands *and thousands* of these "things" into a simulation.

Suppose my combat entity has 1000 objects in it. Furthermore, I want to endow these objects with some behavior they use when some set of sensors is looking for them. Clearly, this behavior should be a function of how many of them there are in the immediate area, so we have an interaction. Marshall and Garrett tell us we need to be more worried about terrain, so these behaviors should also be a function of the terrain (which also probably interacts with the number of objects in a region). It's starting to look like everything interacts with everything, which leaves me with 2^{1000} variants of my behaviors to describe. This would satisfy most people's definition of a big number.

Perhaps the object-oriented modeler will tell me that I really don't have 2^{1000} possible interactions to worry about, as many of the higher-order interactions have negligible effects and this unit of 1000 objects is commanded by a commander object that tells them what to do anyway. Considering just the environmental interactions, that leaves me with perhaps just, say, 2^{20} interactions? That's only a million pieces of data for one unit being looked at by one set of sensors.

This is the bitter lesson of combinatorics, and it completely destroys Marshall and Garrett's notion that I can somehow compute my way out of this box. For the example above, 2^{20} interactions are probably the *fewest* interactions necessary to provide the level of decomposability that an object approach requires. *This is for one unit and one event!* The point is, it doesn't matter how much computing power you have available; we should not decompose things that do not need to be decomposed, because the difficulty of modeling internal interactions so that they mimic the observed aggregate behavior *swamps* the difficulty of modeling aggregate behavior. Consider a simple physical example. Newton's Law of Cooling states that the rate of change of the temperature of an object is proportional to the difference between the object's temperature and that of the surrounding air, and as such is modeled with a first-order differential equation. Now, most physicists would agree that this is a pretty rotten model of what's going on at the molecular level, considering air movements, non-homogeneity of the materials, and so on, but somehow this innocuous equation gives good answers for most situations. If I can get the answer I need using such a simple equation, why bother with the molecules? This leads me to my next point, concerning the regression paradox.

Most *PHALANX* readers have taken a regression course sometime during their careers. A clever statistician teaching regression will reveal to his students that R^2, the coefficient of determination, can never go down if you add more predictors. Aha, say the students. If I can explain more of the total variability with more predictors, I should add a gazillion predictors and drive R^2 to 1. This will get me an A! Of course, the students (and all of us) discover the awful truth sooner or later: not only is wantonly adding predictors not a good thing, it can severely damage the predictive powers of the model. As a result, the best regression model is usually the parsimonious one, and *not* usually the one with the most predictors.

Besides, regressions of most phenomena reveal that there are relatively small sets of predictors that explain most of variability in the response. Why is this? Because a small set of predictors either represents the effects of a multitude of other predictors, or the other possible predictors depend on the behavior of the important predictors. For all of Marshall and Garrett's talk about the need to model at the molecular level, the fact is that few of the molecules act independently.

Another reality intrudes on this party. Given that collecting data for these models costs a *lot* of time and a *lot* of money, I don't want to maintain information on something that has a negligible effect. In every single project I have worked on for the last 16 years, data collection, investigation, recollection, and reduction have consumed the vast majority of the time expended. However, Marshall and Garrett have nothing whatsoever to say about this hidden cost of object models, and this cost is significant.

Finally, Marshall and Garrett speak disparagingly of equations "representing thousands of things," and reject stochastic approaches to modeling information transmission. But, does anyone model without using mathematical relationships? For all the talk about object decomposition, I frequently find object modelers punting to the dreaded equations I would expect them to reject. At a recent MORS meeting, I heard a speaker extolling the virtues of his tremendously detailed and complex combat model, a model that represented platoons, squads, vehicles, and even individual buildings. A lone questioner asked, "How do the engagements get arbitrated?" After a short pause, the answer came back: "Lanchester equations."

Axioms: Theirs and Mine

A useful way to summarize my point is to note that distributed-interactive-object-oriented-plug-and-play literature seems to rely on a small number of axioms. I can only infer these axioms, but they appear to be as follows.

Object Axiom 1: **No mathematical function relating the behaviors of multiple objects can be as good as modeling the objects themselves.**

Object Axiom 2: **Accuracy and explainability increase with the number of objects.**

Object Axiom 3: **Data will be available for any object decomposition, no matter how detailed.**

Clearly, I disagree. I offer the following counter-axioms based on my experience.

Counter-Axiom 1: **There exists a mathematical relation among objects that can be refined to accuracy in a finite and small number of steps, *given the right team of operators and analysts is calibrating it.***

Counter-Axiom 2: **Modeling difficulty increases *exponentially* and explainability goes to *0* as the number of objects increases.**

Counter-Axiom 3: **Data are never immediately available for anything, and their procurement consumes the majority of the available study time *with probability 1.***

The risks inherent in using object-oriented methods seem clear in the face of Marshall and Garrett's overriding assumption (that we need lots of objects to capture the

(See **NUMBERS FROM COMBAT** *p. 27)*

Judgments in Military Research: The Elusive Internal Validity Issue

Clairice T. Veit
RAND

Judgments are used frequently in military operations research and form the basis for a great deal of military policy. Behind this extensive use of judgments lies a well grounded belief that military experts possess knowledge worth tapping as a reflection of truth in their area of expertise, that insight into future decisions can be gleaned from decisionmakers' current judgments, or that judgments are simply the best or only source of useful information. Researchers obtain judgments from subject matter experts (SMEs) to better understand and predict items of military interest covering a broad spectrum of issues such as the location and nature of future conflicts, required readiness levels and how to achieve them, technology and weapon system selection and employment, force structure, and so forth. Such judgments are often analyzed in some fashion and their implications for policy directly interpreted. Judgments are also used extensively to formulate combat simulation algorithms and thus they importantly influence simulation outcomes that serve as a basis for policy decisions. Hence, understanding the validity of claims, inferences, or interpretations based on judgment data underlies their credibility and usefulness in policy arenas.

When it comes to judgments, the predominant validity issue in the minds of most researchers and clients alike is whether the judgments reflect "real world" behavior, hereafter called *external* validity. However, an experimental investigation into the external validity aspects of judgments *depends on* having answered a generally unknown *internal* validity question: "Does the interpretation of judgments accurately reflect the values and thoughts of the respondents?" Not resolving the internal validity question renders the external validity question moot—external validity cannot be appropriately investigated.

Because of its dominant effect on the credibility of research and analysis based on human judgments and the fact that it is commonly ignored or unrecognized, internal validity—what it means, what it affects, the penalty for ignoring it, and how to achieve it—is the main subject of this article. The aim is to enlighten and encourage researchers who choose to base their research on judgments to employ research techniques that provide knowledge about the internal validity of their conclusions.

Validity Criterion

To say that a statement is "valid" implies that it is somehow "true." The criterion proposed here for concluding that a statement is "true" is that the statement has been proposed as an *hypothesis* and tested against and supported by data. The notion of using judgment data to *test* hypotheses is new to most military analysts and in fact, hypothesis testing is notably absent in military analysis when judgments form the research data base. This idea is not new, however, to measurement psychologists who, over the last half century, have developed experimental designs that make such tests possible.

Internal and external validity issues apply regardless of the *type* of response or format for gathering the judgment data—interviews, written responses, electronic actions—or the form of the judgments—quantitative, qualitative, or a written record. To address them, hypotheses need to be formulated and tested.

The SME Issue

A tertiary "validity" issue that has received a great deal of attention in the military community is SME selection. This issue is discussed briefly here because it is often given validity status in the military community where the external validity question popularly focuses on the *qualifications* of the respondents to make correct judgments. The idea seems to be that, if the "correct" SMEs have been selected, then their statements are valid. However resolutions concerning "correct" SMEs lie outside the realm of hypothesis testing and need to be distinguished from issues that deal with the credibility of a researcher's interpretation of SMEs' judgments. It is certainly important to capture the thoughts and ideas of the "right" people. But, who they are is not a testable proposition but a matter of *definition*, and a resolution lies on one's preference or selection criteria, not on hypothesis testing.

Criteria for SME selection could be based on position, rank, experience, recommendations by people who should know, selection by an appointed panel, test scores, and so forth. Whatever criteria are used, some people will probably not agree with them and question if selected SMEs know what they are talking about. However this opinion, as all opinions, should be formulated as an hypothesis and subsequently tested rather than remain a matter of argument. Such tests could be conducted within the hypothesis-testing frameworks described below.

Internal Validity: Testing Hypotheses About SMEs' Values and Thought Processes

The validity issue that is the main focus of this note is referred to here as *internal* validity. The internal validity question asks if researchers' interpretations of judgment data accurately reflect the ideas, values, and thoughts of the SMEs providing the judgments. This question is rarely if ever asked, let alone addressed, by the military analytic community. In fact, when judgments via interviews or surveys form the research data base, policy claims or implications generally consist of repeating or summarizing the contents of the surveys or interviews. Hypotheses are not tested and thus, conclusions cannot be wrong. If conclusions cannot be wrong, one has to ask in what sense they can be right.

The problem is one of controlling variables so that hypothesis testing is possible—a major experimental requirement in all research aiming to draw causal conclusions from data. The basic idea in testing hypotheses about peoples' values and thought processes is to create questions posed to respondents by simultaneously manipulating two or more factors, generally in factorial combinations. Appropriate experimental designs resulting from these

factor manipulations make it possible to test hypotheses about effects of factors on judgment as well as test among hypothesized algebraic models to explain how respondents valued and processed the factor information. Respondents' values come from the model that explains the data; they are the model's least-squares parameter values given the data. If hypothesized models and their best fit paramenter values do not account for the judgment data they are rejected. So in this advanced psychological measurement framework, obtaining respondents' values goes beyond the numbers or answers they provide to the researcher; it rests on determining an algebraic measurement model that has been adequately tested and has received empirical support for its validity.[1-12]

An example will serve to illustrate the ideas of hypothesis testing associated with this measurement approach. Suppose a researcher's interest is to understand what affects intelligence officers' ability to identify enemy maneuver units (e.g., tank battalions) using information provided by intelligence collection systems. Assume the three factors and their factor levels shown below are hypothesized to be key factors affecting the situation assessment process by which intelligence officers identify enemy units:

(1) Coverage: percent of areas of interest covered by the collection system; factor levels: 10%, 40%, 70%, 100%;
(2) Timeliness: how soon the information is available for use after it is collected; factor levels: (5 minutes; 30 minutes; 90 minutes);
(3) Precision: the level of detail provided about observed systems; factor levels: *detection* (something is there); *classification* (tracked and wheeled systems can be distinguished); and *recognition* (distinction by system type, e.g., tank, air defense, personnel carrier).

Further, assume that the measure of performance (MOP) is the percent of enemy units that could be identified in situations specified by combinations of these factor levels. (These three factors and MOP were used in a similar judgment task described in Veit and Callero[13].) Intelligence officers having expertise in performing the real task would be asked to estimate this MOP for each situation presented in an experiment.

Figure 1 illustrates some hypothetical judgment data that might reflect judgments to 36 situations that would have been generated from a 4 (Coverage) x 3 (Timeliness) x 3 (Precision) factorial design. In each panel of Figure 1, judgments of the percent of enemy units that could be identified is plotted on the y-axis as a function of Coverage on the x-axis; a separate curve is for each level of Timeliness, and a separate panel is for the different levels of Precision.

The slopes of the curves in Figure 1 represent the effect of Coverage (x-axis) on judgments; separations between the curves represent the effect of Timeliness, and the change in the shape and position of the curves from panel to panel represents the effect of Precision.

As can be seen, the hypothesis that these factors affect judgments is supported by the data. Also seen in these hypothetical data are two and three-way interactions among these factors in their effects on judgments. For example, in each panel it makes less of a difference how timely information is if there isn't much information than when more information is available (compare the vertical separations among the three curves when Coverage is 10% on the x-axis with when it is 100%). And, this interaction takes on a slightly different form as curves rise, steepen in slope, and spread out from Panel A to Panel C. Thus, these data also support hypotheses about interactions among these three variables. Tradeoffs among the factors on judgments can be seen from the graphs of the data. To advance beyond data description, one can ask the question, "What algebraic measurement model accounts for the independent and interactive effects seen in the data and can provide the values SMEs place on these capabilities?"

Testing Among Algebraic Models to Measure Observed Effects

With a simple factorial design as shown here, tests among algebraic models to explain effects are limited. However, this design serves to illustrate how models can be tested and rejected when wrong. Let us suppose that the widely used expected utility theory was proposed as a candidate model to explain these data. Expected utility theory is an averaging model that can be written:

$$R_{ijk} = m[(W_C S_i + W_T S_j + W_P S_k)/(W_C+W_T+W_P)] + n, \quad (1)$$

where R_{ijk} is the response to the ijk^{th} cell in the 3-way factorial design (responses might be averaged over those SMEs whose data exhibit the same factor effects (e.g., those in Figure 1), S_i, S_j, and S_k are the scale values associated with the i^{th} level of Coverage, the j^{th} level of Timeliness, and the k^{th} level of Precision, respectively (these would be the scale values or measures attributed as SMEs if this model accounted for their data), and W_C, W_T, and W_P are the weights associated with these factors; m and n are linear scaling constants.

If one did not already know that the data shown in Figure 1 violated the predictions of the model of Equation 1, the model

(See **VALIDITY ISSUE,** *p. 33)*

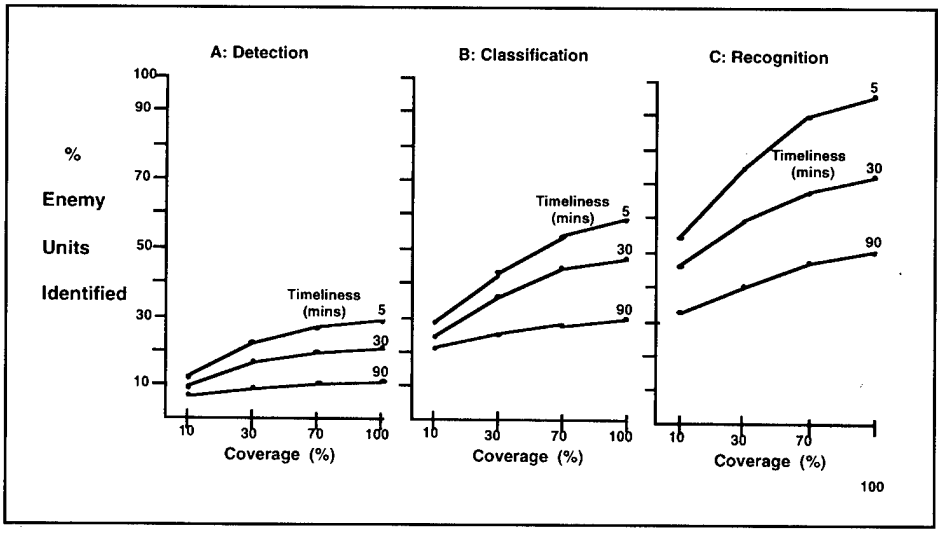


Figure 1: Hypothetical Judgment Data
Observed interactions would lead to rejecting the class of additive models

Some Fallacies in Cost-Risk Analysis

Matthew S. Goldberg
Institute for Defense Analyses

Point estimates of cost for new weapon systems are certain to be incorrect. Interval estimates, which give a range of values with associated probability, provide the decision-maker with much more information. Thus "cost-risk" analysis is supplanting conventional cost analysis in the evaluation of new weapon systems. Unfortunately, cost-risk analysis introduces statistical considerations that are unfamiliar to many cost analysts. As a consequence, several serious, persistent errors have plagued this emerging field. The purpose of this article is to expose a few of these errors in the hopes of inoculating participants in the field. All of the errors discussed in this article were found in papers identified through a Defense Technical Information Center (DTIC) search. However, to minimize embarrassment, none of the authors will be cited by name.

Fallacy #1: Similarly-skewed cost distributions must have non-negative correlation; oppositely-skewed cost distributions must have non-positive correlation.

Program costs are accumulated in a hierarchical system of accounts known as a work breakdown structure (WBS). The cost of each WBS element may be represented as a probability distribution. It has been asserted that two right-skewed cost elements must have positive or zero correlation; two left-skewed cost elements must also have positive or zero correlation; but a right-skewed cost element and a left-skewed cost element must have negative or zero correlation.

It is easy to disprove these assertions. For example, the exponential distribution is right-skewed. There are many families of bivariate exponential distributions for which a negative correlation is admissible. Johnson and Kotz (Chapter 41.3) enumerate the following distributions:

- Gumbel's first bivariate exponential distribution, correlation as low as –.40365;
- Gumbel's second bivariate exponential distribution, correlation as low as –.25;
- Freund's bivariate exponential distribution, correlation as low as –.333.

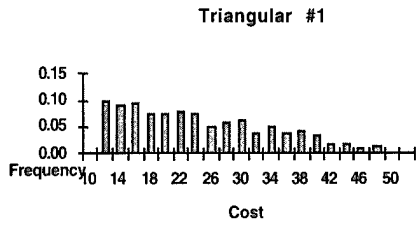
As yet another example, Johnson and Kotz (p. 20) develop a bivariate lognormal distribution by exponentiating the marginal distributions from a bivariate normal distribution. The correlation of the resulting lognormal distribution depends on the variances and correlation of the underlying normal variables. When the normal variables have correlation of –1.0 and respective standard deviations of 0.2, the resulting bivariate lognormal distribution has correlation –.96.

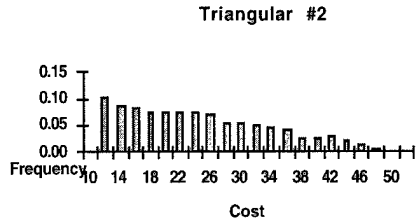
To make these examples even more relevant to cost analysis, I now consider probability distributions specified in triangular form (probably the most common distribution used in cost analysis). The user provides the low, high, and most-likely (i.e., modal) values, and the analyst linearly interpolates the density function between these three points. If the low and most-likely values coincide, then we have a right-skewed right triangle; if the high and most-likely values coincide, then we have a left-skewed right triangle.

I will now provide some empirical counter-examples to the fallacious assertion. Consider two identical, right-skewed triangles each with low value l=10, mode m=10, and high value h=50. Using Microsoft Excel, I first generated 1,000 random draws *independently* from each distribution, so that the (theoretical) correlation between the two distributions is equal to zero. I did so by first drawing random uniform deviates (u), then transforming to triangular deviates (x) using the inverse of the triangular cumulative distribution function:

$$\text{If } 0<u<(m-l)/(h-l), \text{ set } x=l+\sqrt{(h-l)(m-l)u};$$
$$\text{if } (m-l)/(h-l)<u<1, \text{ set } x=h-\sqrt{(h-l)(h-m)(1-u)}.$$

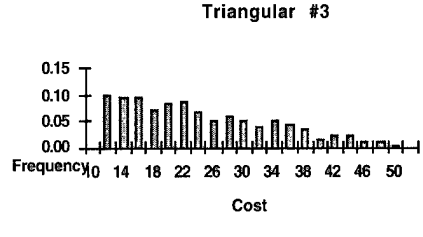
Displayed next are the histograms of the two triangular distributions that I generated, each from a sample of size 1,000. The empirical correlation between the two distributions is equal to –0.004. The empirical correlation differs slightly from the theoretical correlation of 0.0, but this difference would vanish in a large enough sample.

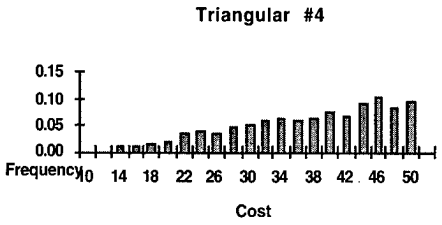




To construct a counter-example, I returned to my Excel spreadsheet. I sorted the 1,000 values of Triangular Variable #1 in *ascending* order, and I sorted the 1,000 values of Triangular Variable #2 in *descending* order. The two histograms remain *exactly the same* as pictured above. But the sorting serves to associate high values of one variable with low values of the other, inducing a negative correlation. In fact, the empirical correlation after sorting is equal to –0.939.

As a second counter-example, I now consider a right-skewed triangle with low value l=10, mode m=10, and high value h=50; and a left-skewed triangle with low value l=10, mode m=50, and high value h=50. Following the same procedure as above, I first generated 1,000 random draws from the two distributions by transforming independent uniform variates. The theoretical correlation is 0.0, and the empirical correlation is 0.029. The two histograms are pictured below.





To construct a counter-example, I returned to my Excel spreadsheet. I sorted the 1,000 values of Triangular Variable #3 in *ascending* order, and I sorted the 1,000 values of Triangular Variable #4 *also in ascending order*. The two histograms remain *exactly the same* as pictured above. But this time, the sorting serves to associate high values of one variable with high values of the other, inducing a positive correlation. In fact, the empirical correlation after sorting is equal to +0.934.

Fallacy #2: If Corr(A,B)>0 and Corr(A,C)>0, then Corr(B,C)>0; if Corr(A,C)<0 and Corr(B,C)<0, then Corr(A,B)>0.

The rationale seems to be that if cost elements B and C each move in the same direction as cost element A, then they must move in the same direction as each other. Conversely, if cost elements A and B each move in the opposite direction from cost element C, then they must move in the same direction as each other. These assertions are sometimes expressed in tabular form:

Case	Corr (A,B)	Corr (A,C)	Corr (B,C)	Possible?
I	+	+	+	Yes
II	+	+	–	No
III	+	–	–	Yes
IV	–	–	–	No

A correlation matrix is logically consistent if it is positive semi-definite (see Marsaglia and Olkin); a necessary and sufficient condition is that the eigenvalues all be non-negative. Here is an example to illustrate that Case II is indeed possible:

Variable	A	B	C
A	1.0	0.25	0.14
B	0.25	1.0	–0.56
C	0.14	–0.56	1.0

The eigenvalues of this matrix are all positive (0.33, 1.10, 1.57), indicating positive definiteness. Positive definiteness may also be tested using principal minors, which are much easier to compute than eigenvalues. One merely checks that the 2x2 and 3x3 sub-determinants (in general, through the *kxk* determinant) anchored in the northwest of the matrix are both positive:

$$\begin{vmatrix} 1.0 & 0.25 \\ 0.25 & 1.0 \end{vmatrix} = .9375 > 0, \quad \begin{vmatrix} 1.0 & 0.25 & 0.14 \\ 0.25 & 1.0 & -0.56 \\ 0.14 & -0.56 & 1.0 \end{vmatrix} = .5651 > 0$$

Here is an example to illustrate that Case IV is also possible:

Variable	A	B	C
A	1.0	–0.12	–0.43
B	–0.12	1.0	–0.61
C	–0.43	–0.61	1.0

The eigenvalues of this matrix are all positive (0.19, 1.11, 1.69), again indicating positive definiteness. Alternatively, both principal minors are also positive:

$$\begin{vmatrix} 1.0 & -0.12 \\ -0.12 & 1.0 \end{vmatrix} = .9856 > 0, \quad \begin{vmatrix} 1.0 & -0.12 & -0.43 \\ -0.12 & 1.0 & -0.61 \\ -0.43 & -0.61 & 1.0 \end{vmatrix} = .3656 > 0$$

Fallacy #3: Maximum negative dependence is achieved by setting all (off-diagonal) correlations equal to –1.0.

When the correlations are unknown, sensitivity analysis can be used to bound the results of a cost-risk analysis. Some rather wide bounds are obtained by considering, respectively, the cases of maximum positive dependence and maximum negative dependence among the cost elements. The former case obtains when every (off-diagonal) correlation is set equal to +1.0. This case leads to the maximum possible variance of the sum of all the cost elements. It has been asserted that the latter case (i.e., maximum negative dependence, implying the minimum possible variance of the sum) obtains when every (off-diagonal) correlation is set equal to –1.0.

In fact, when the off-diagonal correlations in a *kxk* correlation matrix are all equal, they are bounded above by the constant:—1/(*k*–1) [see Johnson and Kotz, p.—51]. Thus for a 3x3 matrix, the negative correlations cannot exceed –0.5 in magnitude; for a 4x4 matrix, they cannot exceed –0.33, etc. As the number of cost elements grows large, a common negative correlation is essentially ruled out, although a few isolated negative correlations may persist.

When these conditions are violated, the correlation matrix is no longer positive semi-definite, and it becomes possible to find linear combinations of cost elements with negative "variance." To illustrate this point, consider a 3x3 matrix with correlations equal to –1.0. The principal minors reveal a failure of positive definiteness:

$$\begin{vmatrix} 1.0 & -1.0 \\ -1.0 & 1.0 \end{vmatrix} = 0.0, \quad \begin{vmatrix} 1.0 & -1.0 & -1.0 \\ -1.0 & 1.0 & -1.0 \\ -1.0 & -1.0 & 1.0 \end{vmatrix} = -4.0$$

To further demonstrate the impossibility of this situation, suppose that the three cost elements each have zero mean and unit variance. Then the sum of the cost elements (i.e., total cost) has *negative variance*:

$$Var(X_1 + X_2 + X_3) = \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} Corr(X) \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix}$$

$$= \begin{pmatrix} 1 & 1 & 1 \end{pmatrix} \begin{pmatrix} 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{pmatrix} \begin{pmatrix} 1 \\ 1 \\ 1 \end{pmatrix} = -3.$$

When the correlations are equal to +1.0, all of the cost elements move in the same direction, reinforcing each other and thereby maximizing the variance of the sum. It might be thought that correlations of –1.0 would have the opposite effect, minimizing the variance of the sum. To see why this is not the case, consider three cost elements A, B, and C. If *Corr*(A,B) = –1.0, then elements A and B move in exactly opposite directions, tending to offset each other and reduce the variance of the sum. Similarly, if *Corr*(B,C) = –1.0, then elements B and C also move in exactly opposite directions, again tending to reduce the variance of the sum. However, this pair of correlations imply that elements A and C move in exactly the *same* direction, so that *Corr*(A,C) cannot be equal to –1.0; in fact, fixing the other two correlations, *Corr*(A,C) must be equal to +1.0. Alternatively, if the three correlations are to have a

(See **COST-RISK,** *p. 12*

(continued from p. 11)

common value, this value cannot exceed –0.5 in magnitude.

Fallacy #4: The formula for aggregating variances is valid only in the case of two-tailed distributions.

Although not restricted to this context, we often think of the variance in terms of symmetric distributions like the normal, for which we can make statements like: "X percent of the probability lies within ±Y standard deviations of the mean." In addition, variances may be aggregated using the following formula:

$$Var(A+B) = Var(A) + Var(B) + 2 \times Cov(A,B)$$
$$= Var(A) + Var(B) + 2 \times Corr(A,B) \times \sqrt{Var(A) \times Var(B)}.$$

If the correlation equals zero, then (taking square roots of the equation above) we find that standard deviations may be aggregated using the "root sum-of-squares" (RSS) formula:

$$\sigma(A+B) = \sqrt{\sigma^2(A) + \sigma^2(B)}$$

These aggregation formulae and, indeed, the very notion of variance were questioned in a paper released by one defense think-tank. The argument is that, once contract awards have been made and budgets have been allocated, costs below the budgeted level will never be realized. Instead, contractors will spend at least their budgets and then some. Thus the *ex ante* symmetric cost distributions are replaced by *ex post* one-tailed distributions. Moreover, according to this argument, the notion of variance and the aggregations formulae are invalidated because there are no longer any negative deviations (i.e., under-runs) to balance the positive deviations (i.e., over-runs). To replace the usual aggregation formulae, the author recommends (without any derivation) the following alternative:

$$\sigma(A+B) = \sigma(A) + \sigma(B)$$

Under this alternative formula, the standard deviations rather than the variances are added.

These assertions are completely false, and the alternative formula is incorrect. The derivation of the conventional (i.e., correct) aggregation formulae in no way relies upon symmetry of the distribution; it requires only that the component variances exist. The source of confusion seems to be as follows. The *ex ante* symmetric cost distributions have larger variance than the *ex post* one-tailed distributions. Aggregating the variances of the *ex ante* distributions (by the conventional formula) indeed overstates the variance of *ex post* total cost. But the solution is not to amend the conventional aggregation formula. Rather, the solution is to apply the conventional formula to the variances of the *ex post* distributions.

Finally, I will now show that the alternative formula, which purports to represent the zero-correlation case, is in fact correct only in the rather opposite case in which the correlation is equal to +1.0. For in that case (and only that case), the conventional

$$\sigma^2(A+B) = \sigma^2(A) + \sigma^2(B) + 2 \times \sigma(A) \times \sigma(B)$$
$$= [\sigma(A) + \sigma(B)]^2,$$

(i.e., correct) formula for $Var(A+B)$ reduces to:

which is equivalent to the alternative formula.

Fallacy #5: The variance of an expression can be reduced by algebraically grouping the terms.

This assertion, which I found in a master's thesis, is by far the most egregious abuse of statistics that I have ever encountered. The author considers the cost of training three people. He assigns an average cost denoted TC (training cost) to each person, so that total cost is equal to:

$$\text{Cost} = TC + TC + TC = 3xTC.$$

He then computes the variance of total cost in two different ways, reaching apparently different results:

$Var(TC + TC + TC) =$

$Var(TC) + Var(TC) + Var(TC) = 3(Var(TC);$

$Var(3(TC) = 9(Var(TC).$

The author then recommends collecting costs on an additive rather than multiplicative basis, because the resulting variance is (apparently) reduced by a factor of three.

This assertion is ludicrous on its face, because one cannot reduce the variance of an expression by simply grouping the terms. The source of confusion is whether the three random variables labeled "TC" (unfortunately, without subscripts) are *always identical in value*, or merely *identically distributed*. If they are identically distributed and uncorrelated, then the first expression for the variance is correct. The second expression is not valid in this case because total cost equals $TC_1 + TC_2 + TC_3$, which cannot be written as 3xTC.

If the three random variables are *identical in value*, then we have:

$$TC_1 + TC_2 + TC_3 = TC + TC + TC = 3xTC.$$

Thus the second expression for the variance is correct. The first expression is incorrect in this case because it ignores the (perfect) correlations among the cost elements {TC, TC, TC}. Incorporating the correlations yields the correct solution:

$Var(TC + TC + TC) = Var(TC) + Var(TC) + Var(TC) + 6(Cov(TC,TC)$

$= Var(TC) + Var(TC) + Var(TC) + 6(Var(TC)$

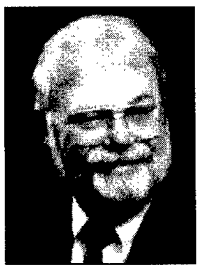
$= 9(Var(TC).$

References

Norman L. Johnson and Samuel Kotz, *Distributions in Statistics: Continuous Multivariate Distributions*, New York: Wiley, 1970.

G. Maraglia and I. Olkin, "Generating Correlation Matrices," *SIAM Journal on Scientific and Statistical Computing*," Vol. 5, 1984, pp. 470-475. ✪

Advanced Distributed Simulations for Analysis (ADSA '96) Workshop



Theodore Bean
MITRE

Introduction. Advanced Distributed Simulations (ADS) and their appropriate use by the defense analytical community were the focus of a recent MORS Workshop, ADSA '96, sponsored by ODUSA (OR),OCNO(N81), HQ USAF/XOM, Joint Staff (J8) and OSD (PA&E). ADSA '96 was another in a long series of MORS Workshops whereby MORS leadership provides a forum for experienced military analysts from all services and the supporting civilian communities to weigh in on a tough question or issue facing the defense community. Generally, these workshops take the form of working sessions where participants strive to clarify the question or issue and propose approaches to assist the DoD in issue resolution. These forums have served the defense community well by providing insight as to where DoD analytical efforts should be focused or leveraged, or, in some cases, to help the analytical community better understand major changes in policy affecting their work.

For this particular Workshop, over 100 senior military and civilian analysts gathered in Williamsburg, VA, from 30 January to 1 February 1996 to have a frank discussion on the applicability/utility of ADS and Distributed Interactive Simulation (DIS) technologies to the analytical community and to document findings and recommendations resulting from the discussion. Much of the discourse that preceded the Workshop and continued through the initial workshop sessions focused on the need for a definition for ADS. **Ed Brady,** the General Chair, believed that any definition would restrict the potential applicability of ADS but finally proffered the following:

> **Advanced Distributed Simulation — The evolving DoD distributed modeling and simulation infrastructure, including synthetic environments, run-time infrastructures, and connected human-in-the loop simulations such as DIS.**

For the purposes of this discourse, both ADS and DIS are distributed but ADS is considered a broader set of modern modeling and simulation capabilities than DIS. For example, ADS has the potential to include the interaction of constructive-to-constructive simulations and constructive-to-live simulations as well as the virtual simulations connected by DIS. Note also that this definition does not require a human-in-the loop as part of any particular instance of an ADS. The variability of humans-in-the loop is a major concern to analysts dependent upon replicability of experiments

Leveraging off lessons learned from previous MORS Workshops, advance preparation for the ADSA '96 included a facilitators training session for chairs and co-chairs in Arlington, VA; scheduled in between the multiple snow storms that blanketed Washington most of this past winter. This training session was to acquaint the chairs and co-chairs with known ADS issues and to prepare them for getting the most out of their respective working groups. In addition, all Workshop participants were provided a healthy reading package in advance of the actual Workshop.

Workshop Charge. During his comments at the opening of ADSA96, Ed Brady emphasized to the workshop participants that they should focus on identifying potential solutions that would be of value to both the analytical community, and the modeling and simulation community. In particular, the participants were asked to identify:

(1) appropriate uses of ADS/DIS capabilities by the analytical community,
(2) inherent limitations and advantages of ADS/DIS capabilities for the analytical community,
(3) current shortcomings of ADS/DIS capabilities that if remedied could significantly improve the utility of ADS/DIS capabilities to the analytical community, and
(4) ways to analyze and test the relevance and quality of ADS/DIS-based tools.

Products resulting from the Workshop were to be focused on clarification of issues and problems, and identification of potential solution spaces. In particular, MORS sponsors' workshop expectations included understanding ways to use ADS-based analysis that differed from those of other analyses as well as identification of classes of analyses most amenable to support by ADS-based tools. And those for which ADS tools are not useful would also be very useful to MORS sponsors. This required the participants to think back on why we analyze what we analyze. Is it because of available analysis tools? Our own limited knowledge? How can ADS-based tools be used to give us more insight? More importantly, giving the changing nature of the questions military analysts are being required to answer, what class of questions looks amenable to analysis by ADS-based tools independent of whether or not they are currently available or technologically feasible? Clearly, an answer to this last question would help focus ongoing research. Another area of particular interest is the identification of experiments designed to assist in the development of measures of merit for analysis of command and control, and other heavily human-related functions. Given the evolving nature of ADS and the changing questions being put to analysts, there was room in the participants' discussions for many differing views of ADS and areas ripe for its potential use.

The remaining ADSA '96 Workshop format consisted of (1) an overview of ADS Technology by Dr. **Randy Garrett** of DARPA, (2) a keynote address by Dr. **Anita K. Jones** Director of Defense Research and Engineering, OSD, (3) a senior practitioners panel, and (4) the kick-off of working group sessions. The working group sessions continued through the last day and concluded with Working Group Chairs reporting out the results of their sessions.

Workshop Synthesis. ADSA '96 Proceedings are being developed that will contain detailed reports from each Working Group. But for the purposes of this article, we will draw upon The Synthesis Panel Report developed by **Stuart Starr** and his panel members. These panel members participated as active members of one of the other six working groups of the Workshop as well as the Synthesis Panel.

(See **ADSA '96,** *p. 14)*

(continued from p. 13)

The nature of the problem. Several of the Working Groups spent considerable time identifying and discussing why the analysis community has been slow to embrace ADS. These issues can be loosely aggregated into two categories: quality and use.

In the area of quality, the Working Groups cited the following concerns among others:

- Ambiguity. Since many of the members of the analysis community are not sure what ADS is, they are not sure how useful it is to them.
- Replicability. Since many examples of ADS feature humans-in-the-loop, there is concern among some analysts about the ability to replicate results. In particular, there is concern about coping with the effects of subject learning and the impact of differences in subject background, experience, proficiency, and morale.
- Credibility. It was acknowledged by several Working Groups that verification, validation, and accreditation (VV&A) of ADS pose particular challenges.

In the area of use, the Working Groups cited the following obstacles among others:

- Cost/resources implications. Several Working Groups observed that it is likely to be relatively expensive to acquire, update, and employ ADS in comparison to more traditional analysis tools. In addition, from a resource perspective, it is often difficult to obtain and train suitable test subjects.
- Time/schedule implications. Numerous Working Groups expressed concern about the extensive time implications of ADS. This includes the time to create an ADS capability, set-up conditions for a specific analysis, perform test runs, and reduce the data that is generated by exercising an ADS. These time considerations can preclude use of ADS when the analyst is given little lead or execution time (e.g., Program Objective Memorandum issue analyses).
- Data implications. Several participants identified a broad spectrum of data issues associated with ADS. These include problems associated with acquiring data needed to set-up and execute ADS; deciding what data to collect and collecting it (note: it may be difficult to introduce ad hoc "hooks" into an ADS to collect desired data); and storing and processing the resulting data (e.g., some of the data of interest may be spoken by the test subjects).
- Efficient, effective exercising of ADS. In order to employ ADS efficiently and effectively, it is necessary to devise and implement experimental designs that provide confidence levels consistent with the issues under study and cope with potential confounding factors that may be unique to ADS (e.g., compensate for potential learning by test subjects).

In view of the many issues associated with ADS, the most natural question is: "Why should analysts bother with ADS?" The Workshop identified several significant opportunities implicit within ADS that suggest that it is well worth the time of the analysis community to explore the application of ADS to a broad range of analyses.

- ADS has the potential to enhance substantially analytic support to a number of application areas.
- ADS may enable analysts to address issues in a richer context than they otherwise could.
- Information technology trends are providing new opportunities and are being enhanced dramatically by the commercial sector.
- The analytic community has the opportunity to take advantage of the extensive ADS investments of other communities (most notably the training community).
- Several senior practitioners postulated that ADS poses the opportunity for analysts to move to "computer assisted analysis" and thereby "return to our roots." They also observed that decision makers are being exposed to ADS through its application by other communities and will come to expect its attractive characteristics.
- **Jack Thorpe,** SAIC, predicted that "In ten years, everything can be connected to everything else and, given the spirit of experimentation that exists within the analytic community, it will!

Findings and recommendations. Based on the insights that the Synthesis Panel gained from participation on Working Groups and Synthesis Panel discussions, several areas were identified that the panel members felt were of overarching importance. For each of these areas, the Synthesis Panel developed findings and recommendations. As these are initial findings and recommendations, they are subject to modification, clarification, and addition as the Proceedings are completed based upon final Working Group Reports.

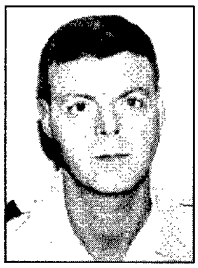
Community Leadership. The Synthesis Panel found that the challenges associated with ADS transcend the abilities and resources of the individual analyst. If analysts are to use ADS effectively, a paradigm shift must occur to enhanced collaborative analysis. This concept, which was put forth by the Analysis and Requirements Working Group, rests on pillars of cross-community shared data, tools (e.g., scenario generation tools, visualization tools), security, Measures of Merit (MoMs), and an integrated family of analysis techniques. The last of these pillars should be emphasized. Most of the Working Groups observed that ADS should not be viewed as a "stand-alone" capability that can be applied to all analytic problems. In general, it must be harmonized and orchestrated with other analysis techniques to compensate for selected ADS features (e.g., extensive time to create and execute) and employed where appropriate. In order to bring about this paradigm shift, key analytic organizations and the analytic community must provide needed stimulation and leadership. A partial listing of recommended actions to be undertaken by the Services, the Joint Staff, and OSD include the following:

- Provide incentives to use ADS.
- Make necessary investments in community infrastructure and accreditation efforts.
- Facilite needed education, training.
- Form ADS teams with the necessary mix of skills. The effective creation and use of ADS demands a team effort.

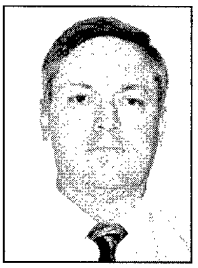
ADS Plan for the Analysis Community. Once the analysis community becomes committed to the enhanced collaborative analysis paradigm, it will need a plan to guide its actions. Historically, the analysis community has been reticent to take this step. For example, several years ago, DMSO sponsored a series of workshops to stimulate the various communities to iden-

(See **ADSA '96,** *p. 21)*

Intelligence Simulation Support to the Joint Training Confederation: Implications for Future Development



MAJ Johnny Bullington
TACSIM
Project Office



Gordon Miller
The MITRE Corporation

In the June 1996 issue of the *PHALANX*, **Anita Zabek** and **Gordon Miller** discussed the Aggregate Level Simulation Protocol (ALSP), a protocol that permits multiple simulations to interoperate, and the Joint Training Confederation (JTC).[1] This discussion included technical interoperability, using the ALSP Infrastructure Software (AIS), and operational or functional interoperability.

This article continues that discussion by focusing on intelligence simulation support to major JTC exercises. This includes a review of the intelligence training audience; the intelligence simulations typically used in major JTC exercises; and several technical architecture issues including security, software interfaces and maintainability; as well as managing future development.

The Joint Training Confederation

A very brief review of the JTC is appropriate to establish background for the discussion of intelligence training. The Joint Training Confederation is an ALSP-supported confederation of simulations. Its purpose is to support Service and joint training of senior commanders and staffs.

Simulations that interoperate with each other through the AIS are members of the JTC. Eight simulations currently directly participate in the JTC as shown in Figure 1

Each of the simulations is supported by a Service or joint activity and provides a representation of the domain of particular interest to that activity. For example, RESA provides the naval surface, subsurface, and air training environment.

Figure 1.
The Joint Training Confederation

The intelligence simulation in the JTC is the Tactical Simulation (TACSIM). This interactive, computer-based training simulation was specifically designed and used to support intelligence training in the Army. In its role in the JTC, TACSIM also supports intelligence training in joint exercises.

The Intelligence Training Audiences

The combat simulations in the JTC primarily support training of senior commanders and their operations staffs (i.e., G3, J3). The combat simulations provide very little training for the general or joint intelligence staffs (G2 or J2) and no training at all for other intelligence activities. Therefore, high resolution, high fidelity intelligence simulations have evolved to meet the specific needs of the intelligence training audiences.

Figure 2 shows the intelligence training audiences and their relationship with the intelligence cycle. Commanders and staffs of intelligence units ensure that intelligence assets are available to support the mission. Collection managers allocate requirements to specific collection assets. Intelligence collectors perform the collection process. Intelligence analysts process raw intelligence and produce intelligence products that are disseminated to commanders and staffs. Commanders and staffs use the intelligence products to support their planning and execution process and generate requirements for further intelligence.

In organizational terms, each command headquarters has an intelligence staff officer, the S2, G2, or J2, and supporting staff.

(See **INTELLIGENCE,** *p. 16)*

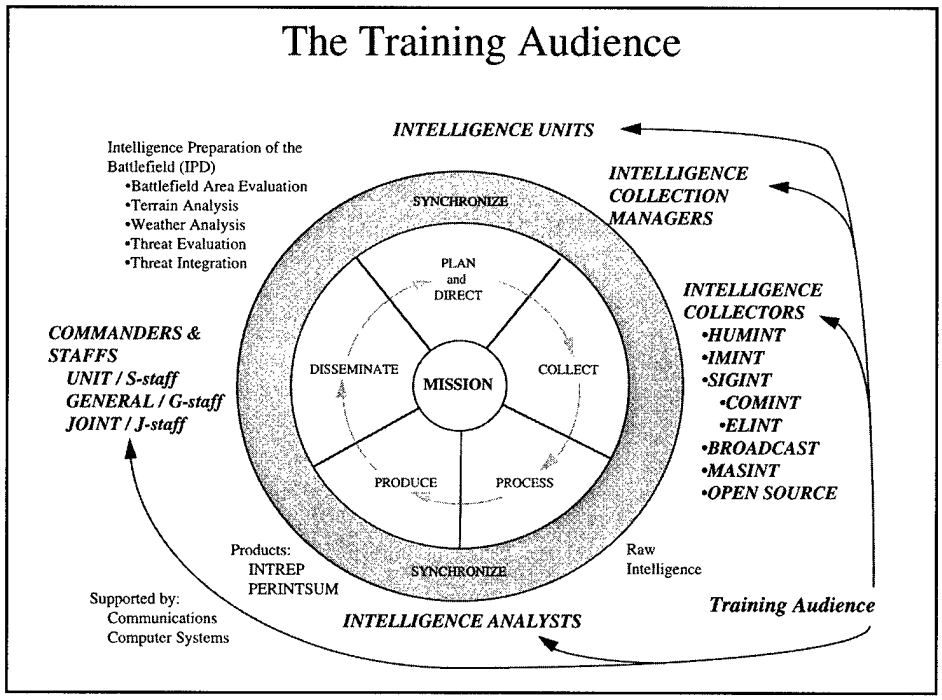


Figure 2.
The Intelligence Training Audience

(continued from p. 15)

The intelligence staff officer advises the commander and other staff sections and coordinates intelligence support for the command. In addition, most major commands have a supporting military intelligence unit such as a military intelligence platoon, company, battalion, or brigade. These intelligence units, with their specialized personnel and equipment, perform the collection management, intelligence collection, and intelligence analysis functions.

The role of the intelligence training simulations is to provide an environment for training the intelligence staffs at the major headquarters and the specialized intelligence functions performed by personnel in the military intelligence units.

Joint Training Confederation Intelligence Architecture

TACSIM is the "official" JTC intelligence simulation. However, to meet the demands of the intelligence training audiences, a more complicated intelligence architecture has emerged over the past few years. Figure 3 shows the major simulations, processes and data flows.

The single block on the JTC architecture diagram labeled TACSIM is really a large collection of simulations and processes which interact to provide a full intelligence portrayal.[2] This collaboration between simulations in the JTC and numerous specialized "peripheral software systems" is typical for simulations in the JTC.

The **data flow** starts with the game truth generated by the combat simulations shown on the left of the diagram. Data flows through ALSP and the CBS Master Interface to the intelligence simulations.

The primary data flow is into the Special Compartmented Information Facility (SCIF) to TACSIM, the National Wargaming System (NWARS), and the Synthetic Imagery Generation System (SIGS). These provide output to the training audience and to the TACSIM Analysis Operations Node (TALON) and the TACSIM After Action Report Utility System (TAARUS).

The second major data flow is through the High Resolution Simulation System (HRSS) to the JSTARS Simulation System (JSS) and the Multiple UAV Simulation Environment (MUSE).

TACSIM is an Army interactive simulation that provides realistic intelligence products for the evaluation of Command, Control, Communications, Intelligence, and Electronic Warfare (C3IEW) functions, systems development, and training.[3]

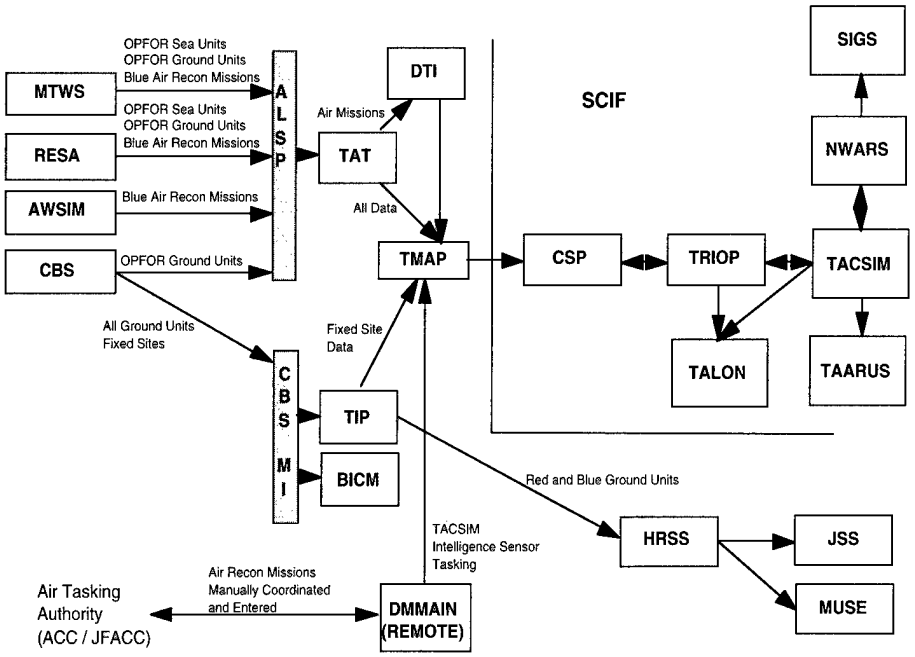


Figure 3.
Typical Exercise Intelligence Architecture

TACSIM:

- Simulates the composition, disposition, and actions of an opposing force.
- Models friendly, or blue force, intelligence collectors to simulate collection tasking requested by participants, players, and testers.
- Produces intelligence reports resulting from simulated collection missions.

TALON is an intelligence fusion process that emulates the function of intelligence analyst cells when these cells are not participating in an exercise. Based on intelligence information produced by TACSIM, TALON can provide summary intelligence reports directly to player intelligence staffs at major headquarters.

TAARUS provides detailed, after action reports. It maintains a detailed history of the intelligence operations during an exercise and can provide this history in a variety of formats suited to intelligence player critiques.

NWARS portrays intelligence collection by national collection assets. This complements the simulation of intelligence collection by theater assets in TACSIM.

SIGS provides hardcopy imagery to players based upon a library of existing imagery and data from NWARS on exercise game truth. Although labor intensive, SIGS adds significant realism to exercises through the interjection of realistic intelligence imagery products.

HRSS disaggregates ground units down to the physical entity level based upon aggregated unit input from the ground combat simulations and an internal rule set.

JSS receives the entity level disposition from HRSS and provides a J-STARS Moving Target Indicator display.

MUSE is a UAV operator training simulator in which a UAV operator flies a UAV over the battlefield and sees a realistic real time video at the operator's control station. Commanders and staffs are beginning to expect these real-time videos as a normal intelligence product.

Architecture Challenges

The technical interaction of the intelligence simulations and some unique problems inherent in them pose numerous challenges for developers and users. Most of the challenges result from the use of simulations to support training needs beyond

that for which they were originally designed.

Strict security requirements add complexity to the system. TACSIM and several of its peripheral software systems operate in a SCIF. The JTC and intelligence simulations outside the SCIF could operate at the unclassified level, though they normally operate at the secret collateral level.

Three processes direct the data flow into and out of the SCIF. The TACSIM Message Automated Processor (TMAP) is the process that collects simulation data and directs the data flow into the SCIF. The Communication Support Processor (CSP) serves as the interface for data transmission between the collateral and compartmented processing environments. The TACSIM Relay for Input/Output Processor (TRIOP) is a message distribution facility that routes messages to and from the CSP.

The need to maintain these processes requires continuing effort and imposes burdens on intelligence simulations as described in the following paragraphs.

Data transfer by file processes are used to implement the above security processes. For example, TAT outputs a data file into a common directory. The next process in the sequence, TMAP, copies this file into its input file directory, reads the file, modifies the file, and then outputs a file to the next process. This is a simple process, but also a potentially slow and I/O bound process.

Simulation data flow is one-way from the JTC into the SCIF. There is no electronic information flow from within the SCIF to the simulations in the JTC because of security requirements. The products from TACSIM and the other intelligence simulations operating in the SCIF are sent to the training audience on real-world intelligence systems.

This causes an automatic requirement for players or controllers to integrate the intelligence with the decision processes in the simulations. There is no electronic connection for intelligence data flow back into the simulations.

Limited simulation time management is another result of the one-way data flow from the JTC to TACSIM. TACSIM can receive the current simulation time from the JTC, so TACSIM time will not get ahead of JTC time. However, TACSIM is unable to send its current time back the JTC, so there is the potential for the JTC to get ahead of TACSIM. In practice, TACSIM typically runs a few minutes behind the JTC, but this is usually not noticeable to the training audience.

Numerous interface programs support the data flow. The number of these programs reflects the evolutionary nature of the JTC and the movement toward greater interoperability between simulations.

The TACSIM Interface Program (TIP) is a CBS Master Interface program and was the first interface between CBS and TACSIM.

As TACSIM became more "joint," the TACSIM ALSP Translator (TAT) was developed to enable coordination of air reconnaissance flights with the air simulations, and collection on non-Army ground units and ships. The DMMAIN TAT Interface (DTI) coordinates air reconnaissance flight information from the TAT with air reconnaissance and sensor information in DMMAIN, a subordinate process of TACSIM.

Each of these processes serves as a conduit from different simulations and the data they convey is not mutually exclusive. For example, Blue and Red unit data from CBS may come through the TIP while Blue unit data from all simulations may come through the TAT. Ground feature information on minefields and similar man-made features may only come through the TIP from CBS at this time.

Multiple tactical intelligence simulations also contribute to the complexity. Unclassified simulations, such as the Battle Command Training Program (BCTP) Intelligence Collection Model (BICM), often perform two exercise functions.

First, Red players should have a realistic and appropriate view of the battlefield in order for operational security planning and execution by the blue players to contribute to the success or failure of blue operations.

Second, an unclassified intelligence program is often convenient in conducting training. Therefore unclassified, low resolution, intelligence programs, such as BICM, are often used.

However, the use of two intelligence simulations performing essentially the same task in major exercises may be an unneeded complexity.

Different operating systems and computer platforms contribute to the software maintenance challenge. The development cycles for the major simulations and the release of new computers and operating systems by commercial manufacturers contribute to the continuing requirements to upgrade the interfaces between systems. As the number of intelligence simulations grows, the difficulty of maintaining the interfaces increases exponentially.

High TACSIM controller requirements result from the initial design of TACSIM and the expanded TACSIM role in supporting training at higher command and staff levels. For example, since TACSIM was designed to support training of intelligence collection managers, there was no provision made for automated collection management for exercises in which collection managers were not playing. Some higher level command and staff exercises do not have collection managers as part of the training audience, so controllers have to be provided to perform this function.

Data base preparation and mirroring remains a significant problem. TACSIM, like many other simulations, has unique data requirements because of its specialized simulation environment. For example, TACSIM requires intelligence signature data for equipment and units in the scenario being played. This data needs to be developed prior to the beginning of an exercise and this leads to duplication of data between simulations. This data base mirroring requires close coordination between the exercise data base developers for each simulation.

Managing for the Future

The TACSIM Project Office approaches its mission of supporting intelligence training with simulations in three ways.

First, TPO will maintain the current intelligence training system. Interfaces will be kept current and TPO will remain responsive to changes required by TACSIM users. This is essential since TACSIM will continue to be the Army intelligence simulation of choice until follow-on systems are developed and deployed.

Second. TPO will address some of the architecture challenges with TACSIM and other current generation intelligence simulations. As resources become more scarce, TPO needs to make some modest investment in improving the maintainability of the current system.

(See **INTELLIGENCE,** *p. 22)*

64th MORSS Wrap Up

Dick Helmuth, SAIC, Program Chair

The 64th MORSS at Fort Leavenworth, Kansas is now in the memory banks, but what wonderful memories those are. The facilities and local support, technical meetings, and social activities all merged flawlessly to provide that ideal setting for personal and professional growth that we have come to expect from the annual MORS Symposium. A very large number of people combined their efforts to make the whole event much greater than the sum of its components and I want to take this final opportunity to acknowledge their contribution.

The cooperation and willingness of Fort Leavenworth to host and support this event cannot be overstated. The local team led by **Phil Kubler** with **Sheila Nickings**, **Dottie Burns**, **Hugo Mayer**, **Carol Mullen** and **Delite Tyler** have set new standards for what can be accomplished. The goal was that every support function flow so smoothly that you don't even notice it, and that level of excellence was achieved. The MORS Office staff, which continues to get better every year, contributed their full share too. **Dick Wiles**, **Natalie Addison**, **Cynthia Kee-LaFreniere**, and **Michael Cronin** (a great program!) deserve all the accolades we can give them, and many more. Several other "regulars" contributed to the smooth functioning of the "office away from home" to include: **Peggy Lambert**, **Bob McIntyre, Gene Schroeder**, **Anne Patenaude**, **Faye Brady**, **Lounell Southard** and a "host" of helpers from the MORS Board of Directors. A special thanks to **Mike Bauman** and the many others who helped to organize the outstanding Western Barbecue at the National Agricultural Center and Hall of Fame. Everyone agreed that it was a true highlight of the week. Also deserving special thanks are **Jean Bauman** and **Marilyn Kubler** who organized and hosted a great Spouse/Guest tour. The final event, a trip by more than 20 to watch the MLS game between the KC Whiz and the DC United organized by Anne Patenaude, was a fitting and successful end to the week.

But, no matter the quality of the support, it is the technical meetings that are at the heart of our symposium. General (ret) **David Maddox** gave an outstanding Keynote address that framed the symposium's theme of "Leveraging Technology for the Military Analyst" perfectly. A special thanks is due to **Stu Starr** and his four session chairs, Dr. **Russ Richards**, **Iris Kameny**, **Nahum Gershon**, and **Marchelle Stahl**, for a superb job of developing and presenting details of that theme in the Special Sessions. Also a special thanks to **Gene Visco** and **Bob Orlov** for organizing an excellent Junior/Senior Analyst Session, and to **Sue Iwanski** for organizing a terrific Poster Session at the Mixer. Also of special note is the great set of tutorials organized by Dr. **Roy Rice**, and the other Special Sessions which made a genuine contribution to excellence.

Everyone should recognize that the real heart of every MORS Symposium is the work done in the Working and Composite Group Sessions. **Kerry Kelley**, as Working Group/Composite Group Coordinator, had the year-long task to make that happen well, and with the help of seven great Composite Group Chairs, **Pat McKenna**, **Bill Mulholland**, **Jay Wilmeth**, **Audree Newman**, **John Green**, **Mary JoAnn Carroll**, and LTC **Mike McGinnis**, 32 Working Group Chairs, and many co-chairs, they made this segment of the program the success it needs to be.

It took a very large number of volunteers to make the 64th MORSS such a success, and only a few of them can be named in an article such as this, but many more deserve our thanks. I know that each recognizes the value of their contribution and how it helped to make the symposium a success for someone else. That is our primary reward. Each of you have a chance to get involved now with the 65th MORSS. The program chair notebook has been passed to Dr. **Harry Thie**, and the Marine Corps at Quantico is beginning preparations already. Don't miss it! ✪

65th MORSS Set for Quantico in '97

Harry Thie, Program Chair

Mark your calendar now—June 10, 11, 12 1997—for the 65th MORSS at Quantico, VA. The theme for the 65th is "Analysis for Complex, Uncertain Times." This theme is appropriate for MORS, for the Marine Corps, and for 1997.

We will be hosted by Marine Corps University. If you have not visited Quantico recently, you will be amazed at all the Marine Corps has underway, and we will take advantage of many of their new facilities while we are there. Quantico is just south of Washington, DC, and we will accommodate daily atttendees via bus service to and from the Pentagon.

You have stated in your responses to our surveys that you want the opportunity at the annual symposium to share expertise among your peers; to meet old and new colleagues; and to learn. We are planning to achieve all of these.

The format of the symposium will be slightly different from last year. The Working Group Chairs asked that more time be allocated to working group sessions so we will do that. We will have 8 working group/composite group sessions during the three days. Thus you will have additional time to engage in peer dialogue. The tutorials on Tuesday, Wednesday, and Thursday will focus on educational opportunities—a chance to learn some new tricks of the trade. And we will schedule social activities for both Tuesday and Wednesday evenings.

There is an energetic group already at work planning and scheduling. Their names appear on the next page. If you have ideas, please contact one of the program staff. Their phone numbers and e-mail addresses are shown.

During the year leading up to Quantico, you will observe some administrative changes in the method and content of the various registration materials. The announcement and call for papers will be sent in October. We will also experiment

65th MORSS Program Staff

Responsibilities	Name(s)	Phone	e-mail
Program Chair	Harry Thie	202-296-5000 x 5379	thie@rand.org
Deputy Chair (Logistics)	Al Brandstein	703-784-3235	brandsteina@mqg-smtp3.usmc.mil
Deputy Chair (Operations)	Kerry Kelley	402-294-8117	kelleyk@j5.stratcom.af.mil
Deputy Chair (Plans)	RADM Pierce Johnson	202-685-5014	radm.johnson@smtpgw.msc.navy.mil
Site Coordinator	COL Richard J. Linhart Jr.	703-784-3235	linhartr@mqg-smtp3.usmc.mil
Plenary/Special Sessions Coordinator	Ted Smyth	301-953-6342	ted.smyth@jhuapl.edu
	Sue Iwanski	516-346-9138	iwanski@grumman.com
	CAPT Lee Dick	703-602-2791	dickl@smtp-gw.spawar.navy.mil
	Anne Patenaude	703-749-5109	anne_patenaude@cpqm.saic.com
Working Group Composite Group Coordinators	Jay Wilmeth	703-695-1763	jwilmeth@is1.js.ml
	Bill Skeith	719-635-2571 x46	bskeith@logicon.com
	Bob Sheldon	703-558-2759	sheldon@brick.saic.com
	LTC Mike McGinnis	914-938-2700	fm0768@exmail.usma.edu
	Mary Bonnet	703-695-5552	bonnet@afsaa.hq.af.mil
Poster Session Coordinators	Ken Martell	703-845-1000	kmartell@calibresys.com
	MAJ Willie McFadden		mcfaddew@tfs.net
	Gene Visco	703-697-1175	viscoep@hqda.army.mil
Tutorials Coordinators	Pat Vye	703-697-8590	pvye@is1.js.mil
	Dean Hartley	423-574-7670	dhx@ornl.gov
	Phil Kubler	913-684-9164	kublerp@trac.army.mil
Best Paper Session Coordinator	Jim Duff	804-498-5646	duff@tecnet1.jcte.jcs.mil
Education Session Coordinator	Yupo Chan	513-255-3362	ychan@afit.af.mil
VIP Coordinator	MAJ Mark Gallagher	402-294-1656	gallaghm@j5.stratcom.af.mil
Junior/Senior Analyst Coordinators	Robert Orlov	703-693-3248	rorlov@is1.js.mil
	Howard Whitley	301-295-1611	whitley@caa.army.mil
	CDR Denny Baer	703-604-0307	baer-dennis@ncca.navy.mil
Spouse/Guest Coordinator	Ginny Wiles	703-751-7290	ginnywiles@aol.com
MORS Staff	Dick Wiles	703-751-7290	morsone@aol.com
	Natalie Addison	703-751-7290	naddison@msis.dmso.mil
	Cynthia Kee-LaFreniere	703-751-7290	clafreni@msis.dmso.mil
	Michael Cronin	703-751-7290	mcronin@tecnet1.jcte.jcs.mil

by sending this electronically to a group of you to see how this will work. If you are interested in participating in this test of electronic registration, please send an e-mail to the MORS office (morsoffice@aol.com), and we will strive to include you. Also, the preliminary program that you will get in January will not duplicate any of the material in the call for papers. This should simplify the communications process. All of the materials will be made available on our Web page at http://www.msosa.mil/groups/mors/. In future issues of *PHALANX*, we will discuss the program schedule in greater detail.

So, mark the calendar: June 10-12, 1997. And, begin to plan on how you will contribute to a great annual symposium. ✪

Brady Wins Wanner Award

Michael Cronin

Edward Brady, FS was named the 1996 recipient of the Vance R. Wanner Award, the Society's most prestigious accolade. The award was formally presented at the plenary session of the 64th MORS Symposium at Fort Leavenworth, KS.

The award is presented annually to a person who best represents the ideals of and has made significant contributions to the operations research community and the Military Operations Research Society. It is named to honor the memory of MORS' first Executive Secretary.

During his career, Brady has contributed significantly to the profession of military operations research and to MORS. For the profession he has served or is serving on numerous advisory boards ranging from studies to intelligence in addition to having been the MITRE Corporations' Washington Group President. For MORS, he served as a MORS President, chaired three workshops, and is the only individual to have chaired two annual symposia. His dedication and demonstrated capabilities clearly make him a deserving recipient of this prestigious award.

Following are Brady's comments upon receiving this award.

"I was very surprised to be nominated for and receive the Wanner Award. Such an honor was the furthest thing from my mind. I very much appreciate and feel strongly that nothing is as personally rewarding as being recognized by your professional peers.

"A little over thirty years ago when I first studied Morse and Kimball at the Naval Academy, I had no idea that throughout my lifetime I would have repeated interesting, challenging and rewarding experiences in conjunction with the analysis of military operations. From my perspective, the current era offers these same opportunities to all of you working in this field. The end of the Cold War and the budgetary driven reductions in the size of our military forces has caused many important questions to be asked by leading decision makers. Many of these questions are very demanding in terms of their need for clarity of thought, openmindness, and attention to appropriate use of methodologies and analytical tools. Some illustrations are: consideration of the likely role and use of the military in the world over the next twenty-five years, the possibility of radically changing operational concepts and use of forces, the net value to the nation of one type of force or weapon compared to another, how to analyze military operations which don't involve combat and attrition, etc.

"Today's Defense establishment is pursuing many interesting approaches to generate the needed answers, although some continue to live in the intellectual past and resist looking at our future. These approaches frequently utilize warfighting experiments, experimental testbeds, and distributed simulation sometimes including large human-in-the-loop exercises. This is a challenging arena for operations analysts. To my mind, our community should be taking a leading role in these areas, but I strongly fear that we are not. Experiments, exercises, and the use of testbeds to support decisions require validity and analytical structure to be valuable in the long run. It is simply not sufficient to witness a few events and decide that "it looks good to me." We must make our voices heard in these arenas, and we can only do so by direct participation. I worry that many of us are not doing so but rather are sitting on the sidelines saying "I'm sorry I don't have an equation or a simulation that fits that question." The end result of such an approach can only be for decision makers to conclude correctly that we are no longer relevant to their problems.

"A brief word about MORS itself. Over the past twelve years, MORS has been like an extended professional family for me, and it can be the same for you. I have learned many things, and met and been befriended by many fine people who are important to me personally, as well as being a valuable resource for me. To have this as an outcome requires more than just coming to MORS, giving a paper and immediately leaving. It takes a commitment of time and effort to participate fully, and thereby to benefit. I urge you to do so."✪

Schnauppner Wins Tisdale Award

Michael Cronin

LT Schauppner receives the Tisdale Award from RADM Marsha J. Evans during the March graduation exercises.

A committee of Naval Postgraduate School Operations Research Department Faculty selected Lieutenant **Craig T. Schnauppner** as the March 1996 recipient of the **Stephen A. Tisdale** Graduate Research Prize. LT Schnauppner earned the award for his thesis titled "An Aircraft Carrier Deployment Scheduling Model." The prize, presented for the best thesis in each graduating class at NPS, honors the memory of LCDR Stephen A. Tisdale, who was killed in a mid-air collision in 1991.

LT Schauppner graduated from the University of California Los Angeles in 1988 with a B.S. in Physics. He received his commission as an Ensign in the U.S. Navy from the NROTC program at UCLA. Previous assignments have included VT-2 and VT-31 at NAS Whiting Field, FL, and VP-31 and VP-9 at NAS Moffet Field, CA. He received a Naval Achievement Medal for superior performance of duty while attached to VP-9. The text of the abstract of his Thesis follows.

Abstract

The Navy's peacetime mission is "to conduct forward presence operations to help shape the strategic environment by deterring conflict, building interoperability, and by responding, as necessary, to fast breaking crises with the demonstration and application of credible combat power." To meet this mission, the Navy deploys aircraft carriers to forward positions throughout the world. A new nuclear powered aircraft carrier costs over 3.4 billion dollars and when deployed carries over 6,000 personnel on board. Considering the cost and man-hours involved in carrier operations, judicious and effective use of these valuable assets is imperative.

The CINCPACFLT Operations Department maintains a five year deployment plan for the six carriers assigned to the Pacific Fleet. Currently, the deployment schedule is produced manually. A feasible five year plan typically takes the carrier scheduling officer one week to generate. This thesis presents an optimization based tool to assist in constructing deployment schedules that maximize the forward presence of Pacific Fleet carriers. The underlying optimization model is different from those in the literature. Instead of using a set covering approach, the problem is formulated as a shortest path problem with side constraints. This formulation allows the problem to be solved more rapidly, thus allowing more possibility for sensitivity and trade-off analyses. ✪

USMA Announces 1996 Hollis Award Recipients

The United States Military Academy Departments of Systems Engineering and Math Sciences announce the winners of the 1996 Hollis Award for excellence in Operations Research:

Cadet **Matthew P. Champion**
Cadet **Jonathan P. Darr**
Cadet **Paul J. Ferrigno**
Cadet **Hernan E. Ruiz**
Cadet **Kyle W. Towns**

Jointly sponsored by the Academy's Operations Research Center and Mathematical Sciences Center of Excellence, the Hollis Award is presented annually to individual cadets, cadet teams, or cadet/faculty teams at the Academy in recognition of excellence in military operations research/systems analysis. The award is presented for military OR/SA work that best contributes to enriching cadet education, enhancing professional development, and linking the Academy to the Army by addressing issues confronting the profession of arms.

The award recognizes the lifetime of selfless service of Mr. **Walter W. Hollis,** FS to the Army and nation. Mr. Hollis currently serves as the Deputy Under Secretary of the Army (Operations Research), a position he has held since 1980.

This year's winning team designed and built a computer simulation tool that enables a user to evaluate the effectiveness of different minefield configurations. The user designs a minefield pattern by placing mines on the screen using a mouse and associated icons. The evaluation program simulates vehicles entering the minefield from different locations and calculates appropriate statistics which can be used to evaluate and optimize the effectiveness of the minefield. The cadet team used this tool to analyze and compare several alternative minefield designs of interest to the project's sponsor, the US Army Armament Research, Development and Engineering Center (ARDEC), Picatinny, New Jersey. The project was directed by LTC **John A. Marin**, Department of Systems Engineering.

The award was presented 30 May in the Thayer Awards Room. Award winners received a plaque, a $50.00 cash prize, a letter of commendation from the Superintendent and the opportunity to present their work at the 64th Military Operations Research Symposium. A permanent plaque engraved with the names of the recipients will be displayed in the Department of Systems Engineering. ✪

ADSA '96

(continued from p. 14)

tify their needs for advanced M&S. While several communities embraced this opportunity (e.g., training, T&E), the analysis community demurred. The Synthesis Panel further observed that there is no long range vision to guide the allocation of limited M&S development resources to support the analyst. Consequently, the analysis community tends to be excessively reactive to near-term needs at the expense of needed longer-term investments.

The Synthesis Panel found that the absence of an "ADS Plan for the Analysis Community" has resulted in fragmented action within the analysis community and limited the leveraging of the resources and expertise that are resident in other communities (e.g., training, acquisition). To ameliorate those shortfalls, the Analysis Council that has been formed under the Executive Committee for M&S (EXCIMS) must take decisive action. It should develop the vision for the next generation of M&S tools for the analyst and draft an associated "ADS Plan for the Analysis Community."

Community Education, Training & Experience. The Synthesis Panel found that a broad set of skills, knowledge, and expertise will be needed by the community if it is to be able to appreciate the capabilities and limitations of ADS and to employ ADS effectively and efficiently. It must be emphasized that the community in question subsumes both analysts and decision makers. In the absence of appropriate education, training, and experience, the results that analysts derive from the application of ADS are likely to be suspect and the decision makers' understanding of the meaning and validity of those results is likely to be limited.

To deal with these issues, the Synthesis Panel put forth several recommendations. First, key Service schools should develop curricula and offer a sub-specialty of "ADS Analyst". Candidate schools include the Naval Postgraduate School (NPS), the Air Force Institute of Technology (AFIT), and the Army Logistics Management College (ALMC). Second, assignments to "ADS Analyst" positions should be of suitable duration and consistency to ensure that necessary experience is acquired and sus-

(See **ADSA '96,** *p. 22)*

ADSA '96

(continued from p. 21)

tained. Thus, for example, once an analyst has qualified as an "ADS Analyst" at one of the above institutions, his/her next assignment should be to an ADS activity or to an organization which may perform collaborative analysis employing ADS. Third, in support of decision maker education and training, appropriate programs on ADS should be developed and offered at institutions such as the Defense Systems Management College (DSMC), the Defense Acquisition University (DAU), the National Defense University (NDU), and Senior Defense Service Schools.

Qualitative assessment of the Workshop. To establish a qualitative measure of the effectiveness of the Workshop, the Synthesis Panel informally assessed the Working Groups against the four ToR objectives, based on its observations of their operations and products. These assessments were in the form of a color (i.e., "green" implied fairly complete achievement of the objective, "amber" implied moderate achievement of the objective, and "red" implied a failure to achieve the objective).

- The first objective was to assess the utility of current and future (3-5 years out) ADS-based capabilities for analytic applications, especially possible areas of new analytic capabilities. The Synthesis Panel rated this as "amber."
- The second objective was to define areas of needed enhancements of ADS-based capabilities to increase their utility to the military analytic community. The Synthesis Panel rated this as "amber."
- The third objective was to increase the military analytic community's understanding of current ADS-based capabilities and near-term improvement programs. The Synthesis Panel rated this as "green".
- The fourth objective was to increase interactions between the military analytic community and developers of ADS-based capabilities. The Synthesis Panel rated this as "green". ✪

INTELLIGENCE

(continued from p. 17)

Third, TPO has initiated the WARSIM Intelligence Module (WIM) project to develop the follow-on system to TACSIM. WIM is an integral component of the Army's Warfighters Simulation (WARSIM) 2000, and will consequently also be in the Joint Simulation System (JSIMS). The architecture issues discussed above will be addressed within those development programs.

Summary

TACSIM, NWARS, SIGS, HRSS, JSS, and MUSE have collaboratively evolved a system to support training of Army and joint commanders and staffs. This system of intelligence simulations is complex and requires constant maintenance.

The system has and will continue to effectively support numerous exercises annually including the Army's Prairie Warrior, U.S. Atlantic Command's Unified Endeavor series, U.S. Forces Korea's Ulchi Focus Lens, among others.

The challenge over the next several years is to ensure a successful transition from the current generation to the next generation of simulations developed under the WARSIM and JSIMS environments. To meet these challenges, TPO will maintain the current TACSIM and TACSIM interfaces to other intelligence simulations, make modest improvements to ensure the maintainability of the current system until the next generation of simulations is fielded, and participate in the development of the next generation of simulations.

References

1 Gordon Miller and Anita Zabek, "The Joint Training Confederation and the Aggregate Level Simulation Protocol," *PHALANX,* The Bulletin of Military Operations Research, Vol. 29 No. 2, June 1996.

2 Marty Brady, Gordon Miller, Bryn Marelli, and William Back, "1996 Joint Training Confederation Intelligence Interface Control Document," MTR 96W000026, The MITRE Corporation, 17 June 1996.

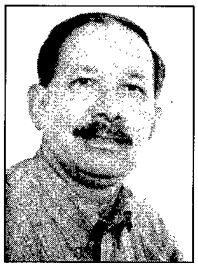
3 Mystech Associates, Inc., "The Warfighter's Guide to TACSIM," Document No. D122-95, TACSIM Project Office, 1 February 1996.

Biographies

MAJ Johnny Bullington *is the Project Director for the US Army Tactical Simulation (TACSIM) Project Office of the US Army Simulation, Training, and Instrumentation Command (STRICOM). Major Bullington has had previous assignments as a Ground Surveillance Radar Platoon Leader and Adjutant/S1 with the 533rd MI Bn (CEWI) in the 3rd Armored Division; Assistant Battalion S3 and HHC Company commander in the 202nd MI Battalion; Collection Manager and Chief, Production in the 513th Military Intelligence Brigade; and Threat System Acquisition Manager at the US Army Combined Arms Center. Major Bullington is a graduate of the United States Army Command and General Staff College and holds a Master of Business Administration from the University of Tennessee.*

Mr. Gordon Miller *is a Member of the Technical Staff in the Information Systems and Technologies Division of The MITRE Corporation and provides systems engineering support to the TACSIM Project Office. Before joining MITRE, he was a Senior Analyst at Science Applications International Corporation and worked on a variety of modeling and simulation and program acquisition projects. Before retiring from the US Army, he had assignments at the HQDA staff, US Army Concepts Analysis Agency, Central Intelligence Agency and the US Army Chemical School as well as numerous tactical units. He received a Bachelors of Science in Chemistry from Worcester Polytechnic Institute and a Master of Science in Business Administration from Boston University.* ✪

Forecasting Issues for the Soldier of the Future



Dr. Oliver Hedgepeth

Military operations research (OR) has a rich history of forecasting or predicting how the warfighter will perform in combat, from the simple back-of-the-envelope model to the million-lines-of-code simulation distributed and linked to other simulations, models and live military exercises. Fine tuning these models and simulations has been the job of major military organizations, such as TRAC, CAA or the schools and academies.

Decades of highly trained officers have come from these schools and academies to tackle the applied problems of military combat and operations other than war. Since the late 1960s, these computerized models and simulations were for the most part quantitative tools and closed systems. Recently, more open systems are beginning to emerge as well as qualitative factors beginning to reemerge into combat and training simulations, which is contributing to a more robust set of OR tools.

However, one area that may be as important is the area of the econometric forecasting model to answer some really tough questions. Recently, many of these questions were, and are, about forecasting technology trends or alternative worlds for the timeframe of Force XXI. Adaptive computer modeling algorithms, better data access methods, and more involvement from senior military leaders has helped improve such forecasts. But, one focus area still needs addressing. That is in the area of forecasting the human element not in the combat model or simulation, but in the real world of recruitment.

To many, econometric models of recruitment supply and demand may not seem as challenging as developing a joint service distributed confederation of simulations. But, consider how well you know that soldier and the key assumptions that describe the soldier's behavior on the battlefield. The M&S community is just beginning to understand the high-low mix of soldiers similar to the high-low mix of weapons used in a model or simulation. But, what does not seem clear is that high-low quality of that soldier, not for this year, but for the years of Force XXI, when many of you will not be in active duty status or may be retired civilians.

For that soldier of Force XXI, what do the officers and enlisted need to know in order to do that task, to carry out their mission, to survive? How does the military take advantage of the information technology to train these soldiers twenty years into the future? These questions may not be as urgent an issue for the would-be officers as it may be with the enlisted. So, the focus shifts to that high school student who graduates in 2020 and decides (somehow) to join the military service. What performance levels will be needed of that green soldier? Will that high school student be able to meet those needs, the military standards? In examining census data, there should be enough supply of high school graduates, but recruiting data show a trend that the demand is not or may not meet the supply. How to address these supply and demand side algorithms depends on the quality of the data and the assumptions the analyst brings to his data and models.

COL **Chuck Kaylor**, US Army Recruiting Command, indicates that the forecasting models for the short and long term seem reasonable. What is missing is a set of mid-range forecasting tools. RAND is also working with OSD and the Army on this issue and is continuing to examine economic trends that may affect the short, mid and long-range modeling and analysis efforts.

For these econometric models, one problem is data. What data is usable? What are the key assumptions about the data that are usable? Are data from the pre-Cold War useful in a post-Cold War world? How do you handle time-series data? Finding a statistically significant answer may require going beyond traditional statistical methods, such as logistics regression into the world of neural networks. The independent variable for these econometric models is usually the number of high-quality students or recruits coming from the high school environment. The dependent variables could be civilian and military pay, employment trends, population size in general or within a recruiting area, race, gender, test scores, military advertising costs and various benefits.

The service academies are an elite intellectual breeding ground for innovations and ideas for alternative analytical solutions to today's complex military problems, beyond those of the combat model and simulation. What we will explore in the future issues are some of these forecasting modeling aspects and key analytical challenges to the OR community. In the meantime, if you have been working on econometric forecasting models, or ethnographic data collections, or related aspects to forecasting human behavior, we would welcome your contribution. Send any proposals, abstracts, model information, analytic results, etc. to email: hedgewo@dcsopspo3.army.mil or facsimile to Dr. Oliver Hedgepeth at (703) 695-9879.

Some suggested readings (courtesy of **Bruce Orvis**, RAND) to acquaint you with this issue are:

- Asch, B.J., "Navy Recruiter Productivity and the Freeman Plan," R-3713-FMP, RAND, June 1990.
- Asch, B.J. and B.R. Orvis, "Recent Recruiting Trends and Their Implications: Preliminary Analysis and Recommendations," MR-549-A/OSD, RAND, 1994.
- Cotterman, R.F., "Forecasting Enlistment Supply: A Time-Series of Cross-Sections Model," R-3252-FMP, RAND, July 1986.
- Orvis, B.R., N. Sastry and L.L. McDonald, "Recent Recruiting Trends and Their Implications: Interim Report," DRR-1175-A/OSD, RAND, August 1995.
- Working, E., "What Do Statistical Demand Curves Show?," Quarterly Journal of Economics, Vol. 41, 1926, pp. 212-235.

Openers

Robert Eberth
Palmer-Eberth

When Dr. **Julian Palmore**, the new Editor of *PHALANX*, asked me to take on the role as department editor of Technical Forum, we found ourselves in immediate and enthusiastic agreement on what the column should be and do. Specifically, the overall purpose of the column will be to identify and illuminate critical technical issues of today in Operations Research in general and Modeling and Simulation in particular.

"Technical" does not necessarily mean "technology." Many of the issues are expected to deal with the theoretic foundation that in fact underlies all that our technologies implement.

The goal may be rather lofty: to stimulate advances in the state of the art, and with that "art" being very broadly defined.

The objectives are to introduce, define, and explore critical issues. The column is intended to challenge the MORS community to look critically at — and engage in a constructive dialogue over — some of our most cherished paradigms as well as at some new or untried alternative technical approaches.

For openers, I want to look at two intertwined issues. The first is the highly popular but increasingly one-sided debate about the "right" kind of simulation approach. The second is the possible dissonance in our pursuit of detail, fidelity, realism, and accuracy.

Type of Simulation

First, *is* there a single type of simulation approach that is inherently superior for particular application environments or for the new generation of major simulations as a whole? The "new generation" I have in mind comprises the Joint Simulation System (JSIMS), initially focused on training; the Joint Warfare Simulation (JWARS), initially focused on analysis and assessment; and the Joint Modeling and Simulation System (J-MASS), initially focused on acquisition and Test and Evaluation (T&E).

The debate to now has centered around two types: (1) aggregate-level or "mass-action" simulations, and (2) entity-level simulations that are "scaled up" to look at results at higher levels of force aggregation. There are other options that are not covered by those two broad categories, and some of the "other" options will be mentioned below, but those two categories dominate the debate today.

Aggregate-Level Simulations

From one perspective, aggregate-level simulations can be viewed as dinosaurs. That is, they were absolutely necessary when computer processing speeds and capacities could not handle the extremely high entity counts and associated computational loads that would be required to scale up from individual entities to theater-level conflicts, even with "low fidelity" entities and deterministic interactions.

That situation has changed dramatically with the capabilities of the computer today, especially if we look at the capacities and speeds of the massively parallel processors. In short, we may no longer *have* to use aggregate-level simulations to model campaign or even theater-level warfare, although large "high fidelity" entity counts coupled with configural considerations and stochastic interactions could fully load even modern computers.

Still, aggregate-level simulations can be highly useful in their own right. They can engender a great deal of insight into a particular decision environment in a relatively short period of time. They have proven highly useful for analysis "at the margin." That is, sensitivity and contingency analyses are employed with the aggregate-level simulation to determine the relative utility of some operational concept or capability.

Their particular advantage, relative to entity-level simulations, is that they are able to very rapidly perform the iterations necessary to those sensitivity and contingency analyses.

Along with that, they have the additional advantage of being "logically transparent" (assuming reasonable documentation). If the analyst using the simulation sees something counterintuitive developing, it is *relatively* easy to determine whether it is the result of some artificiality within the simulation, the result of an error in the simulation or the data, or a new insight into the combat environment or process.

But, by attempting to capture the processes of combat (movement, attrition, etc.) through deterministic algorithms operating on homogeneous masses of forces, such simulations give up the ability to see and work with potentially critical entity-level capabilities and interactions. The possibly unique contributions — both positive and negative — of single entities to the course or outcome of combat are "averaged out" by mass-action models.

Also, structural artificialities within aggregate-level simulations can mean that the abilities of the analyst become the single most critical aspect to achieving valid results. As an example of potential problems at the working level, in aggregate-level simulations the order of processing of actions can take on a crucial and wholly artificial importance with regard to absolute numeric outcomes (e.g., losses to covering fires may be computed prior to losses to mines, giving mines a *structural* disadvantage in terms of their apparent effectiveness as casualty-inflicting weapons).

Similarly, aggregate-level simulations use a wide variety of stylized concepts of massed position and movement (e.g., parallel "pistons" representing land force movement) that can inject discontinuities across artificial geographic boundaries. If the analyst does not recognize and account for those types of *structural effects*, results can be wholly misleading.

Finally, aggregate-level simulations could well be held to be irrelevant for anything other than attrition warfare. In particular, it is difficult to conceive their application in modeling Operations Other Than War (OOTW), where the usefulness of

attrition as a measure of effectiveness would be nonexistent.

The same type of concern applies to their use in modeling maneuver warfare. And, of course, the ultimate condemnation of today: aggregate-level simulations don't look, feel, or act "realistic."

Entity-Level Simulations

Does the above discussion mean that entity-level simulations are then the "correct" choice virtually by default? Certainly their entities can be made to *look* real (particularly if a system were to use the "immersive displays" being marketed today for simulations), and can be made to explicitly represent all the functions of their real-world counterparts.

The potential advantages for that level of detail — for at least the training environment — have to be acknowledged. But does it do anything for the analytic environment? For the acquisition or T&E environments?

Detail, fidelity, realism, accuracy

Entity-level modeling has an almost natural basis of support because it lends itself directly to the never-ending search for greater detail in our simulations.

It's an article of faith: greater detail provides greater fidelity. In fact, the two have become virtually synonymous, both with each other and with "realism."

The draft Glossary of Modeling and Simulation Terms of the Defense Modeling and Simulation Office (DMSO) recognizes three definitions for "fidelity:"

- "The similarity, both physical and functional, between the simulation and that which it simulates."
- "A measure of the realism of the simulation."
- "The degree to which the representation within the simulation is similar to a real world object, feature or condition in a measurable or perceivable manner."

The 24 May 1996 draft of the Operational Requirements Document for the Joint Modeling and Simulation System (J-MASS) also reinforces the conceptual tie between fidelity and detail:

> "[J-MASS shall] Provide for multi-level fidelity/resolution capability. *Fidelity refers to the level of detail* J-MASS and J-MASS models must represent." [emphasis added]

But, is detail for its own sake what we are really pursuing? What are we striving for when we talk about fidelity or realism? Isn"t it accuracy — specifically, the *accuracy of predictions of performance and results* — that we actually are after?

Certainly, for simulations that are intended to support analysis, or acquisition decisions, or T&E, predictive accuracy is the critical attribute of a simulation. That same predictive accuracy may take a back seat to perceptual realism in some narrowly-defined (stimulus-response) training applications, but will be of some concern even there.

If we accept that we are in fact in pursuit of greater predictive accuracy, then the problem for this discussion becomes one of a challenge to the article of faith mentioned earlier: does greater detail at the entity level imbue greater predictive accuracy to the simulation as a whole?

The question seems trivial. How could greater detail *not* promote greater accuracy? The answer is that the two exist and have to be addressed in different frames of reference. The detail that we inject will be in the computer. The accuracy that we seek is in the real world. And, the laws of probability work against us when we transition between those two frames.

For the moment, disregard presentational detail (how real the simulation looks or — for virtual simulations — how real it feels to the simulation operators). The more difficult and germane issues are the degrees of behavioral and logical detail built into a simulation. Two examples help to show what is considered possible in that regard and with current technology.

The Synthetic Theater of War '97 (STOW 97) program is building an ambitious and comprehensive knowledge base of doctrinal behaviors. Conceptually, the synthetic commander will "know" and be able to employ the doctrinally-correct tactical response to changes in his own or the enemy's order of battle, to changes in the environment, and to enemy actions.

As far as I'm aware, STOW 97 will be the high-water mark in behavioral detail in a simulation. At the same time, the draft Technical Requirements Document for JSIMS may indicate that still more is needed, in terms of both behavioral and logical detail:

> [Section 3.1.2.1.3, "Behavior"] "...this behavior [of forces and their commanders] shall include the *ability to 'make mistakes'* and the degradation of unit and human and unit effectiveness due to conditions experienced during military operations. The latter shall include factors such as *combat experience, level of training, fatigue, social, religious, moral* [sic], *environmental conditions (e.g., climate, weather, presence of NBC contaminants), nutrition, sleep deprivation, and political influence.*" [emphasis added]. **1**

There are two problems with the level of detail being sought. First, reality doesn"t exist within the computer. The abstraction of reality that *does* exist within the computer is formed by coded algorithms, control structures, and data handling and transformation routines.

At each level of detail, we must code the cause-and-effect relationships between and among variables. We also must code any cause-and-effect relationships between variables at different levels of detail or scale.

For example, assume we have been able to algorithmically relate the effect of sleep deprivation on a single platoon's combat effectiveness. We then need to code the relationship of that one platoon's effectiveness on the effectiveness of its company. And so on, for as far up the organizational scaling ladder as there is an impact (presumably, "all the way," or the individual platoon's effectiveness didn't matter in the first place).

The problem, obviously, is that to code cause-and-effect relationships, we must first know them and be able to express them in algorithmic form.

To appreciate the full complexity of that problem, consider that none of the degradation-inducing variables in the above JSIMS example are likely to be independent of the others in terms of their combinatorial effects, and that virtually all are certain to be nonlinear in their effects in the real world.

We could, of course, fund a research program specifically to develop and quantify the necessary relationships. It would be a significant program. Or by default, the Government or developer could simply invent the relationships.

As long as the system were to be used purely for training and education, the

(See **OPENERS,** *p. 26)*

(continued from p. 25)

default option could be satisfactory — a panel of subject matter experts (SMEs) could be convened and could develop effectiveness response curves that should ensure that the resulting simulation reflected "roughly right" sensitivity (at least, one hopes, with regard to the *direction* of change in effectiveness) to the behavioral variables. The realities of mutual dependence and nonlinearity would almost certainly be ignored, but might not be significant in light of the fact that the data were to be invented in the first place.

The results would certainly pass a face validation, and could prove highly useful — and appear exceedingly realistic — for training. But, would we want to see the results used for the evaluation of operations plans, or to support acquisition or T&E decisions?

The point is that there are real risks — potentially measured in lives lost — to pushing detail and causal relationships below the level for which we have real data.

The second problem is that if a greater degree of detail is to have any meaning, then the *results* of the simulation must exhibit some sensitivity to the more detailed variables. But if the results are sensitive to highly detailed variables, then a *specific result* may be valid only for the unique case that exhibits the *specific values of the variables that generated that result.*

And — this is where the laws of probability work against us — the probability of observing that same specific set of discrete conditions in a multidimensional, essentially continuous distribution surface is infinitely small — effectively zero.

Again, outside of the training environment there would be real risks in *using* the results of highly detailed simulations, unless we have the luxury of running sufficient replications and sensitivity analyses to ensure that the results are robust.

There is in fact another dimension to the first problem above, but it's one that also points, I think, to a solution. At each successively greater level of detail, we are faced with the problem of having to develop more — and probably more complex — cause-and-effect relationships.

But, no matter how many such relationships we code into a model, we "code out" far more.

The very act of coding ascribes absolute certainty to those relationships in the synthetic abstraction of the world within the computer. Even if the relationships are stated as probabilistic algorithms, they remain certain inside the computer. No other relationships are possible; no others are permitted to exist.

In an absolutely deterministic real world, one in which all cause-and-effect relationships were known with certainty, there would be no theoretic problem. In a highly dynamical, nonlinear, even turbulent system (e.g., warfare), the problem is massive. (To anyone who has doubts that warfare is all three of those, even at the theater level, I strongly recommend Admiral Sandy Woodward's memoirs of the Falklands War, "One Hundred Days.")

As a result, (here I'll go well out on a theoretic limb, and see who cuts it off behind me) I believe there exists a boundary condition on detail (or "granularity") in simulations such that adding detail beyond the boundary is actually counterproductive to predictive accuracy.

I also think it means that "scaling up" entity-level models to answer campaign-level and beyond analytic questions is an approach doomed to failure, simply because much if not most of the turbulence in warfare exists in the *interactions among entities and between levels, and is unpredictable in form, direction, and timing.*

As heuristic evidence, consider that one of the principal characteristics of a chaotic system is *sensitive dependence on initial conditions.* "Initial conditions" is itself a fluid concept; i.e., it can be taken at any point in an event stream, not necessarily at the first point — if that is even known. In Chaos Theory, Lorenz recognized that it was minute variations in initial conditions in a set of *deterministic* equations that caused wildly chaotic behavior in a weather simulation.

In warfare, the examples may not be so neatly stated, but they are legion. As examples of the sensitive dependence on initial conditions in warfare, consider:

- how would the course of war in the Pacific have changed if the Pacific Fleet carriers had been caught along with the battleships at Pearl Harbor?
- how might the course of war in the Pacific have changed if our planes had not found the Japanese carriers at Midway until as little as an hour later (i.e., with planes launched, instead of on deck, fueled, and changing out weapons loads)?
- how might the course (and possibly the outcome) of war have changed in the Falklands if the Argentinean pool of air-launched Exocet missiles — tremendously effective at the engagement level even though mismanaged at the campaign level — been much greater than five?

Shifting gears, can we take anything positive out of all the above? I think so. While it strongly argues *against* the utility of scaling up entity-level models for theater level and beyond analytic applications, I think it also argues *for* entity-level modeling (*stochastic* entity-level modeling) in general to address engagement-level questions.

It principally argues against detail for its own sake in our models. Some general rules of thumb may be useful in:

(1) If there are no hard data to support specific cause-and-effect relationships at some level of detail, either move to a higher level or don't apply the model in any applications *requiring predictive accuracy*

(2) Only specifically include variables that you can in the real world *plan for, observe, or control* (as an example, if my simulation is to support weapons systems acquisition decisions, and I can't plan for or control the religion or politics of my adversaries, those particular variables would be counterproductive in my simulation. Conversely, if my simulation is being used to train Special Operations Forces, and there are unique and tactically useful aspects of the religions of my potential adversaries, *and if I know those aspects and their effects*, then I could be justified in including them.

Aggregate-Level, Again

None of the above means, though, that we must fall back on aggregate-level simulations as the best way to go for campaign-level and above applications. I've helped to build a couple, used several, and

reviewed a few. I've yet to come across *even one* that I could say was predictively valid or accurate in any absolute sense. That means that "accuracy" was absolutely dependent on the skills, knowledge, *and objectivity* of the analysts.

The track record has not been encouraging in that regard. Along with all the problems mentioned earlier with respect to the higher levels of warfare, must go the potential, particularly prevalent in aggregate-level simulations, for subjective interference precluding predictive accuracy.

One example should suffice. A couple of years ago, in a theater-level assessment, I observed a surprisingly high level of attrition of enemy air assets in an extremely short time interval, against essentially no friendly air-to-air losses.

On closer examination, it turned out that the friendlies were engaging the enemy aircraft at maximum weapons range, given detection, and that there was a Blue advantage in detection range capability. Rules of Engagement? What Rules of Engagement? ROEs were not modeled explicitly, and so were ignored completely.

The correct answer, and one easily implemented in aggregate-level simulations, would have been to "turn down" the range capability of the missiles to reflect realistic ROE restrictions.

Wrong! As I was told, the "correct" answer would have so degraded the effectiveness of the friendly missiles that the entire value — and thus programmatic future — of a set of AAW capabilities could have been threatened.

Is it any wonder that we have trouble convincing anyone of the operational value of some C^3 capabilities? Much of their value, of course, is due to their synergistic effect with our weapons systems. That value can only be fully demonstrated (in attrition models) by accurately showing how much more *poorly* the weapons systems perform in the absence of particular C^3 capabilities. As stated earlier, the potential for subjective "tuning" is particularly high in aggregate-level simulations.

Solution?

I think we need to take a much different approach to campaign-level and above modeling, again if predictive accuracy is the goal. And the result may not look at all "realistic."

Imagine that I want to build a model of the U. S. economy, one that would tell me whether I should invest in specific stocks, particular types of mutual funds, bonds, or commodities. I could, given enough research funds, build a detailed entity-level (consumer-level) simulation of the economy (and maybe the rest of the world's economies too, because of the interdependencies). Entity-level population models could be integrated into the whole.

I could then watch resources migrate into certain economic sectors relative to others as time went on. It could certainly *look* realistic. I could watch Dick queue up at the grocery store, watch Jane decide on a Ford rather than a Chevy based on some market statistic (next iteration she might take the Chevy, or a Harley).

No matter how detailed I made it, does anyone believe it would be *predictively accurate*? How about if we tried it at the aggregate level? Blocks of consumers could interact with blocks of goods and services...

Fortunately, in economic modeling, *we've got data.* Reams of data. Which means we can use the tools of statistics and the computer to make sense out of the data — *to let the data dictate the model.*

No, econometrics is far from the ideal answer, even in applications where the data are sufficient. Econometric models carry heavy baggage, not the least of which are rigid assumptions about data and error distributions. But I'm convinced the basic concept is right for campaign-level applications and above; *we've got to let the data drive and shape the models.*

We also don't know yet how to do that in a relatively data-sparse environment such as warfare. We need research into that particular field (Econometric Modeling in a Data-Sparse Environment?).

Much of the discussion here also has pointed up the apparently strong analogies between warfare and turbulence, warfare and "chaos." We need *serious, in depth* research into Chaos Theory applied to warfare modeling. I am personally convinced that real-world warfare results will be found to be much better explained and understood — even modeled — by the nonlinear mathematics of Lorenz, Huberman, Mandell et al, rather than by the determinism characterized by Lanchester. Please let us know your thoughts.

Robert Eberth
13371 Meadowsweet Drive
Fairfax, VA 22033-1108
eberthr@smtp-gw.spawar.navy.mil

Footnotes

[1] As an aside, when I first saw this list of "soft factors" to be included in the simulation, I was struck by the absence of the soft factor that is arguably the most important of all. I'm indebted to the late Colonel **Trevor DuPuy,** USA, who stated at a MORS More Operational Realism in the Modeling of Combat (MORIMOC) working group several years ago, "the most pervasive element influencing performance in combat is *fear in a lethal environment.*" If you've been there, you know he's right. It's also the most radically nonlinear element I can imagine. ✪

NUMBERS FROM COMBAT

(continued from p. 7)

necessary detail. We don't know whether all these objects have important effects, so we have a choice. We can aggregate their behavior via another object to avoid getting our combinatorial lunch eaten, in which case we would have been better off using the body of aggregate methods we know and only decomposing as necessary. Or, we can spend a great deal of time and money gathering data on these objects only to discover we either can't model the interactions or we've run out of time and money - in which case we've accomplished nothing. Tell me again why we're doing this?

I may be hopelessly old-fashioned, but this is what I think Operations Analysts are supposed to do: we should build abstractions of reality, check that they make sense for attainable limiting cases, refine them for regions of interest, and use them to support decision making. In Marshall and Garrett's world, I pick a method, decompose my model to the lowest possible level, gather an extraordinary amount of data, and hope that the *secondary* issues of whether my model is correct will sort themselves out. I believe that object-oriented approaches have their place, and I suspect I will employ these methods sometime during my career. But objects are just another hammer, and I don't use hammers on things

(See **NUMBERS FROM COMBAT** *p. 28)*

NUMBERS FROM COMBAT
(continued from p. 27)

that don't look like nails.

Editor's Note: I welcome Major Yost as the first independent contributor to this column. His article forcefully raises both practical and theoretical concerns about a growing body of effort in the OR community. As such, his presentation fits nicely into one of the column's purposes: inquiry into the inherent capacity of various OR approaches for fidelity to the real world they purport to represent. The debate over "object-orientation" (which is every bit as fundamental as Yost suggests) seems to lead to another inquiry: into our judgments as to *What* out there in the real world in fact constitute the *"phenomena"* and their characteristics to be represented. Such judgments determine whether and how we focus our analytic attention from the first instance (i.e., before *any* representations are attempted): on "objects," "relationships," "variables" or "things" and their causal interactions, or whatever. Such judgments usually either originate essentially with us or derive from a given OR approach's theoretical structure. Inquiry into the character and empirical supportability of the phenomena as judged (whether originally or derivatively) is another of the column's purposes. Reader comment is invited on either topic of inquiry, of course including Kirk's argument. ✪

MORS PRESIDENT

(continued from p. 3)

"Foundations" theme discussed above and the slate of Special Meetings for the next year that leads us into the Quantico Symposium.

Our Special Meetings Committee will continue under Dr **Stuart Starr's** leadership this year and has an ambitious slate of proposed meeting topics and work underway for meetings right around the corner — this fall and winter. Please feel that you are an integral part of both the 65th MORSS and the Special Meetings this year and GET INVOLVED!! The MORS Officers and Board desire to focus "outside" on the conduct of quality meetings and publications this year. We are moving quickly under **Jackie Henningsen's** Joint Analysis SAG to get a "Quick Response Analysis" meeting in place for the week of 1 October 1996, a most ambitious plan!! More detail on the complete list of Special Meetings in planning for this year are contained in the VP(MO) article "VEEPS PEEP" on page 5.

The Sponsor's and Past Presidents Luncheon Meetings conducted during the MORSS provided us with some valuable insights into what is HOT for the next year. One suggestion is to start working on an International Symposium for participation by selected allies to be conducted perhaps every other year. That meeting could grow to be of the size and quality of our annual Symposium. Across the board we heard a consistent concern with meaningful analysis to support the new joint, coalition and interoperability environments — as well as Operations Other than War. There is concern that we can provide the right tools for collaborative analysis and decision making under uncertainty. Watch more for these topics as we unfold our special meetings agenda and prepare for Quantico in '97.

The following table lists the new officers and their phone and email addresses. Don't hesitate to call me directly or any other member of our Board of Directors and management team. We welcome your input and active involvement in this, your Military Operations Research Society. The new MORS Organization is listed in the following diagram. Get to know your VP's and Committee Chairs and let us hear from you.

I consider it to be one of the most important events in my professional life to have been elected to serve for the next year as President of your Society. It is an honor and privilege to join the ranks of the illustrious former office holders of this position. I'm looking forward to a great MORS year together! ✪

MORS Executive Council

Fred Hartman
President
202-298-7166
hartmanfe@aol.com

Jerry Kotchka
Vice President, Finance & Management
314-232-2284
jkotchka@mdc.com

Kerry Kelley
Vice President, Meeting Operations
402-294-8117
kelleyk@j5.stratcom.af.mil

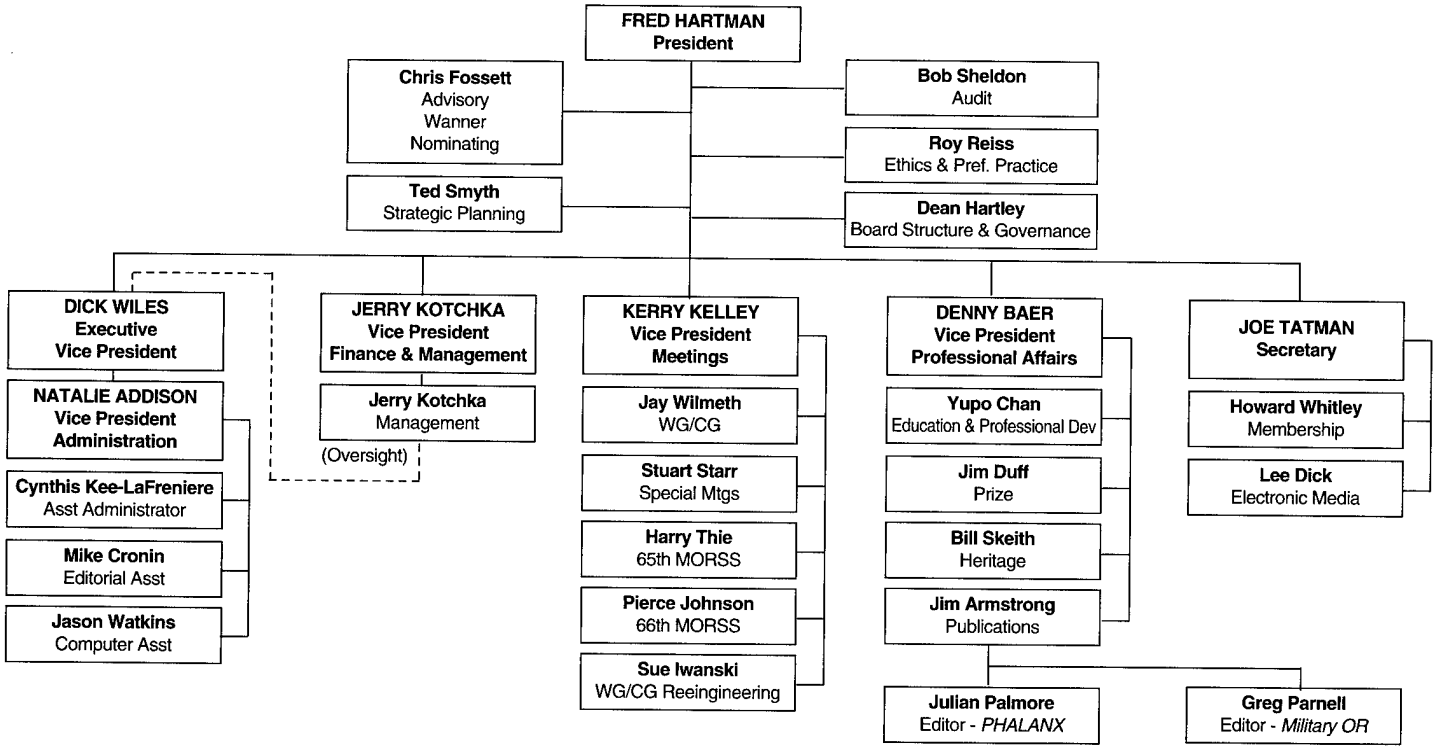
CDR Dennis Baer
Vice President, Professional Affairs
703-604-0307
baer-dennis@ncca.navy.mil

Joe Tatman
Secretary
703-558-2788
cjtatman@brick.saic.com

Christine Fossett
Past President
(202) 512-2956
cfossett@msis.dmso.mil

Dick Wiles
Executive Vice President
703-751-7290
morsone@aol.com

MORS 1996-97 Organization



starting point and catalyst for a wide-ranging discussion of V&V issues and activities by the entire Colloquium. The remainder of this article highlights the discussions involving several common threads that ran throughout the conference. In general:

- Both the DoD and non-DoD M&S communities share similar objectives, e.g., credible M&S;
- Both the DoD and non-DoD M&S communities are experiencing similar problems, e.g., data;
- The tools and methodologies used by both communities (excluding DIS and ADS) are fairly equivalent.

Data and Test Sufficiency

Almost universally, presenters indicated that their modeling or V&V efforts were hampered by two interrelated problems: the lack of both sufficient (real-world systems) testing and data. In recent years there has been a reduction in testing due to many factors such as reduced budgets, increased test costs, safety considerations, range constraints, etc. This has led to a corresponding loss of data collection opportunities afforded by the test process. (Even when testing is available, there are two instances—fidelity mismatches between model and test data, and corruption of data via the data collection process—where the test data is later found to be unusable for V&V purposes. It is not always possible to determine *a priori* if certain test data will fall into one of these categories.)

Besides reductions in testing, other reasons for this dearth of quality data—similar to that experienced within DoD—include:

- *The data for a particular problem domain is too difficult to collect or does not exist.* This was illustrated in Dr. Kleinberger's presentation of Human Anatomic Finite Element Models, when he indicated that, in certain test scenarios, it was infeasible to obtain and use live human or animal subjects, and that surrogate subjects (crash dummies or human cadavers) were incapable of providing realistic data.
- *The data is available, but the collection method corrupts the data being collected.* An example—a sensor inserted into a flow field would itself disrupt the flow field it's trying to characterize. This disruption may alter the test data enough to make it unusable for validating certain high-fidelity model applications.
- *The data is available but is inappropriate for use in model validation efforts.* Mr. Hicks noted, in his presentation of dispersion models, that actual low-altitude atmospheric wind patterns are both extremely random and localized. This localized data is easiest to collect even though wide area wind patterns are generally of prime interest. The issue: is it possible, or appropriate, to validate a model when there is a mismatch between the data and model fidelities (in this case, using detailed localized wind data to validate a more aggregated, wide area wind model.)
- *The data is available but is considered too sensitive for release outside of the owning organization due to public policy, legal, corporate proprietary, or similar considerations.* These situations were observed in Dr. Knepell's presentation on modeling AIDS transmissivity (public policy) and Mr. Neat's rivet crack propagation presentation (where the aerospace industry was reluctant to share its proprietary models or data because of possible use against them in future litigation).
- *Data sets provided from multiple authoritative and pseudo authoritative sources contain data voids, conflicts, or inconsistencies.* The opposite situation also applies: data production sources may refuse release of these problem datasets—even with appropriate caveats—due to potential data misuse or damage to the organization's reputation once the data are beyond their control.

Face Validation

One result of this lack of data is the heavy reliance on expert face validation and peer review as means of establishing model credibility. Colloquium participants felt these techniques could be made more effective if they were repeatedly accomplished over time. Reviewers bring to the model evaluation task both their subject matter expertise and some level of familiarity with the strengths and weaknesses of the best models currently used within their problem domain. Repeated successful face validation of a model for different aspects of the problem domain by such review panels would raise the model's credibility within that community.

While widely accepted, Colloquium participants noted that these approaches have their own limitations:

- They assume you can make an adequate observation of the phenomenon being modeled. However, the difficulty in observation is, at times, the cited justification for using M&S (to provide insights into the phenomenon being observed or modeled.)
- They assume that there are accepted community standards against which these models can be judged. However, it was noted that there is intense debate within many disciplines as to the theoretical constructs that correctly portray the behavior of their problem domain. Such debate would also extend to the review of models based on these constructs.

One method to possibly offset the observability problem is through the use of appropriate visualization tools. These tools would permit the observer to see what is happening with the model over time. This would benefit the analysts that are using these models to gain better understanding of very fine nuances within their problem domains. (Later, these tools would be particularly effective in building the accreditation authority's confidence in the model's credibility.) One note of caution: the analyst must know where or what in the model to visualize. The tool can provide misleading information if the visual is set incorrectly.

It was also suggested that practitioners not automatically limit themselves to face validation if they are not able to compare model output with sufficient instrumented live test data. For example, one method to increase model confidence is by stimulating the model in such a manner that its outputs can be analytically solved and then compare calculated versus model outputs. Another means of establishing a model's goodness could be through use of Dr. Knepell's experimen-

tal design methodology. (It was noted that more work was needed to develop better measures of goodness, as statistical point estimates such as bias, variance, and correlation are not adequate for this task. It was suggested that fuzzy logic techniques be used as a means of determining model goodness metrics.) Often overlooked, the sensitivity analysis is one of the most useful tools of the analyst trying to validate a model. The bottom line is that there are other available techniques, and practitioners should make good use of them.

V&V of Model Hierarchies

One of the lessons learned presented during Mr. Gordon's discussion of rail car wheel heat treatment simulations was that no one model does it all. This illustrates a trend within many M&S communities over the past several years towards the use of model suites or toolkits to accomplish problem domain analysis. The composition of these suites support problem analysis vertically (using models that address the problem domain in increasing levels of fidelity), horizontally (using models of similar fidelity that address different aspects of the problem domain), or cross-cutting (that is, for an analysis spanning several problem domains.)

Within these suites, outputs from one model are often used as input to another model(s). Each model will produce data at a certain level of error. Data transformations and interfaces facilitating data flow between the models will increase this error. This raises a question: how should practitioners conduct V&V of integrated modeling suites, and how would it differ from V&V of stand alone models? The Colloquium participants concluded that V&V should be conducted in two steps: first, conduct V&V on individual models within the suite, then conduct V&V on the model suite in the configuration used during actual analysis (treat it as a black box model.) This is an approach similar to that recommended for the V&V of DIS environments.

Colloquium members also discussed the linkage (if any) between V&V activities and accreditation (that is, meeting the decision-maker's needs). Many Colloquium members appeared to support the position that the needs of the decision makers must be the primary guide when planning V&V activities. This would drive a need to interactively develop requirements and acceptance criteria with the decision maker, then validate to that criteria

Legacy Models and the V&V Process

Legacy models, as defined by DoD, are models developed in the past which are still in use that were not implemented using today's standards. (Some legacy models have been modified to meet the current standards, thereby extending their usefulness and interoperability with newer, standards based models.) Even though these models that have been used successfully for many years in a given problem domain, they must nonetheless be subjected to the VV&A process when they are used for new purposes. This was one of the lessons learned highlighted during Mr. Keltch's Tertiary Oil Recovery Information System (TORIS) presentation. The TORIS modeling suite had been used successfully in analyzing petroleum-related problems at the national level. In a clear departure from previous usage, TORIS was used in a royalty reduction study aimed at increasing oil production (and thus increasing the in-ground reserves' royalty producing capability) within Wyoming. Initial model results were at variance with the anticipated output. Subsequent model evaluation (which greatly benefited from the use of proprietary petroleum industry data) resulted in several changes that addressed the resolution problems and corrected the model output. Policy decisions made as a result of the TORIS-supported study substantially raised the value of these in-ground reserves.

Cost of V&V Activities

Model V&V and IV&V processes are expensive propositions. The Colloquium participants focused on the issue of what should be spent on V&V. It was the general consensus of the participants that V&V costs are situation-dependent, based on such factors as study objectives, model complexity and the V&V level of effort (LOE) needed to reach a predetermined confidence level. Two LOE methods were mentioned:

- The Criticality Analysis and Risk Assessment (CARA) method used by NASA that rates each software function against seven weighted risk and criticality drivers—is one LOE (but not cost) estimation tool. (CARA is a *software* IV&V tool—it makes no provision to the V&V of the underlying conceptual model, so it would probably underestimate LOE required for full model V&V.)
- Colloquium participants that are active in the distributed interactive simulation (DIS) VV&A arena noted that a DIS costing tool capable of providing cost estimates—as a percentage of DIS development costs—for a user-specified DIS V&V LOE is currently under development. An extension of this work—to develop a generic costing process—is also part of the DMSO VV&A initiative.

Budgeting for V&V presents an interesting program management problem. In order to provide V&V funding during program start-up, you must somehow accurately link (estimated) V&V and overall software costs. One of our Colloquium presenters suggested that, because of the variable nature of these requirements, V&V should not be estimated as a fixed percentage of the software development costs. Program managers would be hesitant to agree to a *variable* funding scheme for model V&V. However, if over time you could demonstrate the effectiveness of V&V to these managers, through the use of well-documented success stories, they might be more inclined to approve front-loading a V&V funding wedge. As an example: Mr. Gordon's presentation on shakedown stress estimation studies of rail car wheels showed that, by modifying the parameters used to determine when to turn down these wheels, one could save $875,000 *annually* at a one time combined cost (model development, V&V, and study implementation) of $700,000. While acknowledging this argument's effectiveness, it has thus far proved difficult to locate very many of these success stories.

V&V — Personnel Dimensions

Any discussion of modeling activities will eventually touch upon how the

(See **VV&A** *p. 32)*

human being impacts the process (or, as stated in the September 1995 edition of the *PHALANX*, **It's the ANALYST, Stupid!**). From personal experience, Colloquium participants have found that in conferences where this subject is discussed, the analyst is always rated as far more important than modeling tools and data to the successful completion of a study. (It was previously mentioned that successful validation using peer review or face validation was heavily dependent on the quality of the participants.) Two questions were raised at the Colloquium's wrap-up session:

- What type of people, in terms of skill sets, should be considered when putting together a V&V effort?
- How do you track and tap into this expert base? Or, more importantly, how do you identify those organizations with the appropriate expertise that can provide a user with credible model V&V?

There was no consensus answer *per* se identified for either of these questions. The discussion on model validation (as used by DoD) pointed towards a team effort that includes expertise in statistical methodology and analysis, software development, and in the functional area being modeled. The latter point—the criticality of competent subject matter expertise (SME)—was part of the lessons learned drawn from the TORIS and rail car heat treatment case studies. It should also be noted that Colloquium presenters, by and large, were highly experienced subject matter experts within their problem domains. The (software) verification step—the subject of two Colloquium presentations—should be relatively easier and more straight forward, than the validation step. This is primarily due to the availability of better verification tools (such as Computer Aided Software Engineering tools, Configuration Management tools, and critical path analysis) and techniques (Dr. Miller's presentation on software indicated they uncovered 153 specific technique *classes* during their literature review).

The second question is of particular interest to the DoD participants (especially M&S users). There are no known plans to establish dedicated V&V provider organizations within DoD, and many DoD organizations do not possess an in-house V&V capability. Currently, the ability to locate and assemble the appropriate V&V team is a function of a program's available resources and knowledge its personnel possess on both the appropriate functional and modeling communities. (Editor's note: since the Colloquium, there has been work within DMSO to provide accreditation authorities with access to M&S technical support services.)

Standardized Terminology

Early on in the proceedings, it became clear that, while Colloquium members shared the same understanding of the activities that constitute model V&V, they did not share a common terminology. This required the Colloquium members to cross-map their definitions with those of the DoD standard definitions of verification, validation, and accreditation. It was the consensus among Colloquium participants, at the conclusion of the conference, that words must have precise meaning within the modeling community to avoid potential confusion among its members.

Parting Comments

The most perfect V&V methodology, flawlessly executed, is wasted effort unless it ultimately increases the decisionmaker's confidence in the model—and *contributes* to their willingness to implement the model-supported end product. (This *is not* meant to downplay the critical impact that the analyst wields regarding the decision maker's acceptance of the end product.) This was the case with our presenters:

- Public policy changes in several Third World countries resulted from Dr. Knepell's study on AIDS transmissivity;
- The frequency of aircraft inspections were influenced by Mr. Neat's crack propagation analysis;
- The maintenance of rail car wheels and railroad track was influenced by the stress analyses described by Mr. Gordon;
- The value of Wyoming's in-ground oil reserves royalty producing potential was increased because of Mr. Keltch's TORIS model analysis.

As one Colloquium participant noted during final wrap-up discussions, "I didn't come here expecting to hear the Golden Nugget, and I didn't hear it—but I got a lot of benefit from the Colloquium." One of the benefits was a confirmation, based on the cases cited by our presenters, that V&V can contribute to the organization's bottom line mission. In the final analysis, that's what it's all about.

Biographies

Dr. Patricia Sanders *is the Deputy Director, Test, Systems Engineering and Evaluation for Test Facilities and Resources in the Office of the Under Secretary of Defense (Acquisition and Technology). In this capacity she is responsible for oversight of the Department of Defense's Major Range and Test Facility Base and the development of test resources such as instrumentation, aerial targets and other threat simulators. The MRTPB comprises more than fifty percent of the DoD land resources, represents a capital investment of more than $25 billion with an annual investment of three billion dollars, and employs approximately 54,000 government and contractor personnel. Dr. Sanders is also the Department's focal point for the use of modeling and simulation in the development and test of weapons systems and the Chair of the Acquisition Council for Modeling and Simulation.*

Ray Miller *works distributed interactive simulation and verification, validation, and accreditation issues for the Technical Support Division of the HQ USAF Directorate for Modeling, Simulation, and Analysis. He is a retired Air Force officer with an extensive background in airlift and air mobility operations. He received his undergraduate degree in mechanical and aeronautical engineering from the University of Pittsburgh, a graduate degree in systems management from the University of Southern California, and is currently completing a graduate degree in information systems and software systems engineering from George Mason University.* ✪

Research on the Causes of Dynamical Instability in Combat Models by **Julian I. Palmore** is available as USACERL Technical Report 96/95.

Abstract

Combat models are nonlinear deterministic models of decisionmaking processes that deal with attrition of opposing forces. They contain dynamical instabilities that destroy the robustness of a simulation and interfere with a simulation's consistency, fidelity, portability, reliability, and validity. Nonlinear behavior in a model's design is a primary source of instability. Striking effects of instabilities on the performance of a combat model are found in computer arithmetic that cause global divergences of simulation runs, branchings on thresholds set by decision tables that can be triggered inadvertently by small errors in the computation of state variables, and structurally unstable decision logic.

Difficulties in reducing and eliminating structure instability and structural variance arise from the use of modeling paradigms. Several paradigms are distinguished: discrete event simulations, embedded dynamical systems, and models of computer arithmetic. A dynamical systems viewpoint illuminates properties of a model and its computer simulation that are essential for verification and validation. By viewing computational processes as dynamical systems that are embedded in discrete event simulation, fundamental difficulties for models of the real world are brought to light.

This report discusses examples of unstable behavior, demonstrates ways to reduce or eliminate sources of instability, and suggests strategies for designing valid combat models that are consistent, robust, and stable.

VALIDITY ISSUE

(continued from p. 9)

could be tested as follows. First, the model's least-squares parameters would be estimated from the data (analogous to estimating regression weights, except that here all parameters are simultaneously estimated from the data). This can be done using a least-squares program such as Chandler[14]. These estimates are then computed in the model to produce a predicted point to coincide with each data point shown in Figure 1. For Equation 1, the resulting predicted curves would be *parallel* in each panel, indicating the independence among factors predicted by this model. Since the data are clearly interactive, the expected utility model *and its best-fit parameter values* would be *rejected* as being representative of how SMEs process and value this factor information in making their judgments. It is this ability to reject hypotheses that lends credibility to conclusions about factor effects and models and SMEs' values to account for those effects.

Tests of model predictions against data are more powerful when data are graphed as in Figure 1. Although statistical tests are informative for assessing significance of observed effects, they say nothing about what data look like. Interactions can take many forms; different forms lead to different conclusions about tradeoffs among factors, algebraic models, and SMEs' values to account for the effects. Statistical goodness-of-fit indices that correlate predicted with obtained data points can be misleading for assessing a model's credibility, because these indices can be quite high for incorrect models[15,16] and even higher for an incorrect than a correct model.[17] An incorrect model would lead to incorrect conclusions about SMEs' values and tradeoffs. Graphic displays (such as shown in Figure 1) of predicted *and* obtained data allow assessment of location, magnitude, and direction of model/data discrepancies and thus guidance for selecting a model that could explain the data.

Once a model has received empirical support for its validity, it can be used to predict events along the entire continuum for factors that have physical values such as Timeliness and Coverage. This is possible because the function relating the values used in the experiment to the model's scale values (S values in Equation 1) can be estimated (the resulting functional relationship is referred to as the Psychophysical Function in Psychology). Model values along the continuum can be estimated from this function and computed in the model to get the model's predictions.

The importance of a good experimental-design/model match for model testing can be pointed out with respect to the factorial design used in this example and Equation 1. Had the model in Equation 1 been represented as a simple additive model with no weights, $R_{ijk} = m (S_i + S_j + S_k) + n$, the marginal means of the factorial design would have been the least-squares values of the factor levels (the S values)[18]. However, for a model that multiplies weights and scale values as in Equation 1, a factorial design as that illustrated in Figure 1 does not permit separate estimates of these two parameters; they are confounded in the marginal means. Additional experimental design features are required to unconfound these two parameters so that weights and scale values can be uniquely estimated. (See the algebraic logic behind this in Veit et al.[16].)

For the data in Figure 1, a multiplicative model of the form $R_{ijk} = m (S_i S_j S_k) + n$, may be a rewarding model to test since it predicts the observed divergent interactions and changes in interactions shown from Panel A to Panel C. (See Veit and Callero[13] for a description of the models that had to be rejected for a similar task, and a description of the geometric averaging model that explained features of the judgment data other models, including the expected utility and multiplicative theories, failed to predict.)

It can be seen that, as questions are omitted (i.e., data points in Figure 1 are erased), the structure imposed on the data by the factorial design begins to disintegrate, as does one's ability to assess the effects of factors on judgments or to determine a model to explain those effects.

External Validity

The second validity issue is referred to here as *external validity* because it concerns the relationship between SME's values and thoughts to external "real-world" events. This validity question can be stated as follows: "Do SMEs accurately perceive what *their* or others' decisions or behaviors would be in specified situations?"

(See **VALIDITY ISSUE,** *p. 34)*

VALIDITY ISSUE

(continued from p. 33)

When using judgments, external validity can be thought of as a second order level of validity. Only when it is understood what and how factors impact judgments is it possible to explore the match between those judgments and analogous "real world" outcomes by manipulating the factors known to affect SME's judgments in the defined "real world" environment. (The term "real world" is placed in quotes because the validity research medium selected for the external validation test must permit *manipulation* of the hypothesized factors; the real world naturally confounds them, thus precluding tests of their causal effects.) Some in the military community have understandably voiced a great deal of concern over the external validity issue, but few to no satisfactory investigations of external validity have been forthcoming.

An external validity test of the data shown in Figure 1 could proceed as follows. Let us suppose that a field exercise was selected as the external validity medium for this test. The three factors—Coverage, Timeliness, and Precision—would be *manipulated* in the same 4 x 3 x 3 factorial design displayed by the data points in Figure 1. Manipulation requires that factor levels are made to occur and not occur. Thus, conditions would have to be set so that the level of precision *resulted in* a detection, classification, or recognition of the enemy's vehicles as specified by the particular cell in the factorial design (or a surrogate would have to be used); this is also the case for Coverage and Timeliness. If all 36 conditions were not practically possible to perform (perhaps because it would be prohibitively expensive), the end points of each factor could be selected, thus reducing the number of trail cells to 8. Details of the experiment would have to be worked out so that learning problems that occur using people in exercises is avoided.

In any case, when data on the percent of enemy force elements identified are collected under each different condition comprising the experimental trials, these data can be compared in structure and magnitude with the judgments. A superimposed plot of the exercise data on the same graph as the judgment data will reveal the location, magnitude and direction of exercise/judgment deviations. These kinds of validation methods are illustrated in Chapter 4 of Callero, et al.[19], for validating results from a distributed interactive simulation.

The economic burden of using live exercises as an external validity medium is perhaps their greatest drawback. And, despite the "realistic" setting, people express concern about the artificial environment of field experiments as they do with judgment experiments. In field experiments, participants do not experience real combat. Rather, they pretend to fight an enemy on all too familiar terrain (e.g., the National Training Center) in a relatively benign environment. A further concern is the inability to investigate the effects of proposed capabilities on performance of systems not yet fielded; restrictions are to systems that exist and can be procured for the experiment. Primarily however, because of the tremendous economic burden of field experiments, researchers turn to other research mediums to address their research questions. Primary among those research mediums are simulations and SMEs' judgments. Each medium has its associated validity issues.

Comments

If there is to be a military science, assertion-based research (i.e., opinions and claims in the absence of supporting data) and descriptive research need to be placed in their proper perspective—as information bases for hypothesis formulation. Research methods to test hypotheses need to be adopted, regardless of the research medium—judgments, simulations, live exercises. The well-known research guideline for addressing causal questions in any research area, whether it be medicine, education, physics, astronomy, biology, and so on, is to control for variables that could provide alternative explanations of observed results. Experimental control is especially important when policy implications resulting from research endeavors are costly — either in terms of money or lives.

A major purpose of this article is to suggest to researchers who *choose* military judgment as a basis for their research, to use research methods that place claims or policy implications emanating from judgment data in a framework that allows them to undergo validation tests, that is, allows them to be *rejected* when wrong. If SMEs' judgments are to be used, it is important to be able to appropriately interpret their meaning.

References

1. Anderson, N. H., Functional measurement and psychophysical judgment, *Psychological Review*, 1970, 77, 153 - 170.
2. Anderson, Norman, H., Algebraic rules and psychological measurement, *American Scientist*, 67, 1979, 555 - 563.
3. Anderson, N. H. *Foundations of Information Integration Theory,* Academic Press, 1981.
4. Birnbaum, M. H. The nonadditivity of personality impressions, *Journal of Experimental Psychology Monograph*, 1974, 102, 543-561.
5. Birnbuam, M. H. Comparison of two theories of "ratio" and "difference" judgments, *Journal of Experimental Psychology: General,* 1980, 109, 304-319.
6. Birnbaum, M. H., and Sutton, S. E. Scale convergence and utility measurement. *Organizational Behavior and Human Decision Processess,* 1992, 52, 183-215.
7. Birnbaum, M. H. and C. T. Veit, Scale convergence as a criterion for rescaling: information integration with difference, ratio, and averaging tasks, *Perception and Psychophysics,* 1974a, 15, 7-15.
8. Birnbaum, M. H. and C. T. Veit, Scale-free tests of an averaging model for the size-weight illusion, *Perception & Psychophysics*, 1974b, 16, 276-282.
9. Krantz, D. H., R. D. Luce, P. Suppes, and A. Tversky, *Foundations of Measurement,* Academic Press, New York, 1971.
10. Krantz, D. H. and A. Tversky, Conjoint-measurement analysis of composition rules in psychology, *Psychological Review,* 1971, 78, 151-169.
11. Luce, R. D. Where Does subjective expected utility fail descriptively? *Journal of risk and Uncertainty,* 1992, 5, 5-27.
12. Veit, C. T. Ratio and subtractive processes in psychophysical judgment, *Journal of Experimental Psychology:*

General, 1978, 107, 1, 81-107.

13. Veit, C. T. and Callero, M. A Method for Measuring the Value of Scout/Reconnaissance, Santa Monica, CA: RAND, MR-476-A, 1995.
14. Chandler, J. P., STEPIT—Finds local minima of a smooth function of several parameters, *Behavioral Science,* 14, 1969, 81-82.
15. Anderson, N.H., Integration Theory and Attitude Change, *Psychological Review*, 1971, 78, 171-206.
16. Yntema, D. B. and Torgerson, W. S., Man-Computer Cooperation in Decisions Requiring Common Sense, *IRE Transactions on Human Factors in Electronics,* 1961, HFE-2, 20-26.
17. Birnbaum, M. H. The Devil rides again: Correlation as an index of fit. *Psychological Bulletin,* 1973, 79, 239-242.
18. Veit, C. T., Callero, M, and Rose, M., Introduction to the Subjective Transfer Function Approach to Analyzing systems, Santa Monica, CA: RAND, R3021-AF, 1984.
19. Callero, M., Veit, C. T., Gritton, E. C., Steeb, R., Enhancing Weapon System Analysis: Issues and Procedures for Integrating a Research and Development Simulator with a Distributed Simulation Network, Santa Monica, CA: RAND, MR-340-ARPA, 1994.

Biography

Clairice T. Veit is a senior research psychologist at RAND, Santa Monica, where she has worked on military judgment and decisionmaking problems over the last 17 years. She developed a research framework, The Subjective Transfer Function Approach, for testing hypotheses about how outcomes are affected in complex systems such as command and control where multiple groups of subject matter experts perform interdependent activities. She has also developed research frameworks for assessing validity issues inherent in combat simulations and distributed interactive simulations. She has applied these methods extensively in her military operations research. She received her Ph.D. in Measurement Psychology from UCLA in 1974. ✪

PHALANX is actively soliciting articles. If you are interested in publishing in *PHALANX,* or if you would like a copy of the *Instructions to PHALANX Authors,* contact Michael Cronin in the MORS office at (703) 751-7290; morsoffice@aol.com.

VEEP'S PEEP
(continued from p. 5)

Reengineering Committee

We have established an ad hoc committee this year to review the current structure of MORS Working Groups and Composite Groups. Since the last comprehensive review of the MORS WG/CG structure in 1993, incremental changes have since been suggested and adopted. However, much has changed in the defense and national security community over the last few years, and these changes may affect the types of operations research analyses needed to support decision-making. Therefore, over the next months we will be reviewing the functional structure that organizes our members and our annual symposium. The Chair of this committee, **Sue Iwanski**, has firmly taken hold of the objectives to review and either validate current WG/CG structure with minor adjustments or to prepare alternative structures. According to the committee's plan of action and list of milestones, during September she is compiling comments from the membership to assist in the review of the current structure. If you have suggestions or ideas, please contact Sue.

In this review of recent and planned activities for the meeting operations for MORS, I hope you agree that we are dedicated to enhancing the quality and usefulness of military operations research. We are looking forward to a challenging and rewarding year! We appreciate your suggestions and involvement – the following names, numbers and e-mail addresses are provided for easy reference. ✪

VP(MO)	Kerry Kelley, USSTRATCOM/J502 kelleyk@j5.stratcom.af.mil	(402) 294-8117
65th MORSS	Dr. Harry Thie, RAND thie@rand.org	(202) 296-5000 x5379
WG/CG Coordinator	Jay Wilmeth, SETA, Inc. jwilmeth@seta.com	(703) 695-1763
Special Meetings	Dr. Stuart Starr, MITRE starr@mitre.org	(703) 883-5494
Reengineering	Sue Iwanski, Northrop Grumman iwanski@grumman.com	(516) 346-9138
66th MORSS	RADM Pierce Johnson radm.johnson@smtpgw.msc.navy.mil	(202) 685-5014

Military Operations Research Society
101 S. Whiting Street, Suite 202
Alexandria, VA 22304-3483
(703) 751-7290
FAX (703) 751-8171
E-mail: morsoffice@aol.com

INFORMS
901 Elkridge Landing Road
Suite 400
Linthicum, MD 21090
(410) 850-0300 (800) 446-3676
FAX (410) 684-2963
E-mail: informs@jhuvms.hcf.jhu.edu

THE LAST WORD

An Abundance of Articles

Dr. Julian Palmore

The *PHALANX* queue has been growing steadily all year. After the exciting 64th Military Operations Research Society Symposium at Fort Leavenworth the queue grew by an additional 10 articles to over two dozen. What this means for everyone, especially department editors and authors, is *PHALANX* is well-recognized as a good place to publish to reach the Military Operations Research community. With the abundance of articles submitted for consideration for publication come inevitable challenges and opportunities. The main challenge is to publish articles of interest to the *PHALANX* audience that are important and timely. As a bulletin, *PHALANX* must be immediately useful to its readers. In each issue of 36 pages, there is room for 12-18 pages of articles. An article's average length is about 3-4 pages. This means there may be as few as four or as many as six articles in an issue. The *PHALANX* Department structure has been rejuvenated with a total of twelve departments, fully functioning. Submitted articles may appear either within a department or stand alone. The implication of a large queue of long articles is there will be fewer appearing in each issue than with shorter articles. So please consider the length of an article as you prepare it for submission to *PHALANX*. In our case, "**shorter is better**." ✪

CHECK OUT MORS' NEW HOME PAGE!

http://www.msosa.mil.inter.net/mors/